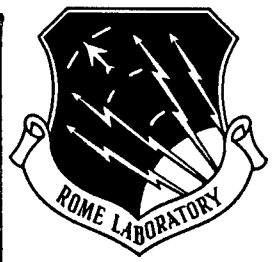
RL-TR-95-164, Vol IV (of five)
Final Technical Report
September 1995



HIGH-LEVEL ADAPTIVE SIGNAL PROCESSING ARCHITECTURE WITH APPLICATIONS TO RADAR NON-GAUSSIAN CLUTTER, The Problem of Weak Signal Detection

University of Massachusetts at Amherst

Prakash R. Chakravarthi (Syracuse University)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-95-164, Vol IV (of five), has been reviewed and is approved for publication.

APPROVED: Vincent C. Vannicola

DR. VINCENT C. VANNICOLA
Project Engineer

FOR THE COMMANDER: Donald W. Hanson

DONALD W. HANSON
Director of Surveillance & Photonics

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (OCSS) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1995	Final Apr 91 - Jun 94

4. TITLE AND SUBTITLE
HIGH-LEVEL ADAPTIVE SIGNAL PROCESSING ARCHITECTURE WITH APPLICATIONS TO RADAR NON-GAUSSIAN CLUTTER, The Problem of Weak Signal Detection

5. FUNDING NUMBERS
C - F30602-91-C-0038
PE - 62702F
PR - 4506
TA - 11
WU - 1B

6. AUTHOR(S)
Prakash R. Chakravarthi (Syracuse University)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Massachusetts at Amherst
Department of Computer Science
Lederle Graduate Research Center
Amherst MA 01003

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (OCSS)
26 Electronic Pky
Griffiss AFB NY 13441-4514

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RL-TR-95-164, Vol IV (of five)

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Dr. Vincent C. Vannicola/OCSS/(315) 330-2861

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

We describe the accomplishments of a collaborative effort carried out by a team of researchers from the University of Massachusetts at Amherst (headed by V. Lesser), Boston University (headed by H. Nawab), and Syracuse University (headed by D. Weiner) on the development of a high-level signal processing architecture called IPUS (Integrated Processing and Understanding of Signals). Based on the IPUS generic testbed architecture and "radar and sound understanding" (RESUN) architectures, we have been able to transfer IPUS technology from a LISP environment to a C++ environment for use in an IPUS Radar Clutter Analysis Testbed. Though not as well-developed as the sound understanding application because of its newness, this radar testbed has still clearly demonstrated the potential of IPUS-like technologies for CFAR processing of radar returns. There has also been significant development of knowledge for weak signal detection. This knowledge has involved the application of the Ozturk algorithm to hypothesize the distribution of data in a clutter patch based on a small amount of data. Also, techniques have been developed for partitioning the radar surveillance volume into background noise and clutter patches, for weak signal detection in K-distributed clutter, and the efficient use of Rejection Theorem for Weibull clutter generation. Though the effort on the application of IPUS to communications was given less priority, we still did some interesting theoretical work on (see reverse)

14. SUBJECT TERMS
Radar, Signal processing, Artificial intelligence detection. Clutter, Non-Gaussian

15. NUMBER OF PAGES
162

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

13. (Cont'd)

weak signal detection in communication systems subject to spherically invariant random processes (SIRP) interference.

Abstract

This investigation is motivated by the problem of weak signal detection in a strong clutter background. The concept of the Locally optimum Detector has been used to address this problem. The problem of weak signal detection has been extensively addressed in the literature when the received radar samples can be modeled as independent and identically distributed. However, this issue has not received much attention when the received radar samples are correlated and have a non-Gaussian probability density function. Also, performance analysis is not generally carried out for finite sample sizes.

This thesis addresses the performance of Locally Optimum Detectors in radar weak signal detection for finite sample sizes where the radar disturbance is modeled as a correlated non-Gaussian random process. The theory of Spherically Invariant Random Process is used for statistical characterization of non-Gaussian radar clutter. In particular, the K-distribution and the student-T distributions have been considered as models for radar clutter. A canonical form is established for the Locally Optimum Detector that is a product of the Gaussian linear receiver and a zero memory nonlinearity. The functional form of the zero memory nonlinearity depends on the approximation used for the underlying radar clutter probability density function. Since the weak signal detector is nonlinear, thresholds for specified false alarm probability cannot be established in closed form. Given a specified false alarm probability a new method for threshold estimation based on extreme value theory is derived that reduces by orders of magnitude the computation and sample size required to set the threshold. Once the threshold is set the performance of the Locally Optimum Detector is carried out for finite sample sizes through computer simulations. Finally, the concept of the Amplitude Dependent Locally Optimum Detector is introduced which has significantly improved performance over the Locally Optimum Detector for K-distributed clutter. The performance evaluation of the Amplitude Dependent Locally Optimum Detector is also carried out for finite sample sizes.

Acknowledgments

I would like to thank Professor Donald D. Weiner for guiding me through research with stimulating ideas and a keen attention to details. To my parents, I am indebted, for encouraging me in the pursuit of higher education from very early on.

This work was supported by Rome Laboratory, U.S. Air Force, under contract numbers SCEE-DACS-P35498-2 and UM 91S025/28140, which I gratefully acknowledge. Thanks are also due to Michael Wicks, Russel Brown, Lisa Slaski and Jim Michels of Rome Laboratory for the discussions that we had during the course of this research. I am also grateful to Dr. Rajan Srinivasan and Dr. Aydin Ozturk for helping with ideas and resources on subjects related to this work. To Dr. Harry Schwarzlander, I owe a big thanks for reading this manuscript thoroughly. I would also like to thank Murali Rangaswamy and Mohamed Slamani on whom I sounded many ideas and received critical feedback. Finally, I would like to thank all others who made this effort possible in various ways.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Weak Signal Problem

In radar applications it is found that the received target signal is contaminated with clutter and thermal noise. The received signal due to undesired reflections from land, sea, atmosphere etc. is called *clutter*. The thermal noise, which is generated by the receiver hardware, is typically modeled as a Gaussian random process. This kind of noise is always present. Depending upon the situation, the clutter may or may not be modeled as a Gaussian random process. Also, the power associated with the background clutter may be orders of magnitude larger than the receiver thermal noise or the desired signal power.

In modern radars, temporal and spatial processing are used to separate the target from the clutter. For example, the received signal from a target having a radial velocity with respect to the radar will experience a Doppler shift. If the target spectrum appears in the tail of the clutter spectrum, then conventional frequency domain techniques can be used to extract the target from the clutter. Similarly, if the spatial spectrum of the target does not overlap that of the clutter, performance will be limited by the background noise rather than the clutter. In this research use is also made of temporal and spatial processing. However, we are interested in the case where the target temporal and spatial spectra cannot be separated from the strong clutter. By definition, this is referred to as the weak signal detection problem. Given a Range-Doppler-Azimuth cell in which a target is to be detected, it is assumed that the signal is larger than the background noise but much smaller than the clutter.

In the weak signal problem the performance is limited by the clutter even after

temporal and spatial processing. Therefore, it becomes very important to identify the clutter plus noise probability density function. This density function is the N^{th} order joint density function of the received radar samples $r_1, r_2, ..., r_N$ in the absence of a target signal. The received waveform can be modeled as a random process. Since we will be sampling this process at N time instants, we need to know the N^{th} order joint probability density function (PDF) of the N random variables. In this research effort the performance measures of radar receivers are analyzed, given the N^{th} order PDF associated with the random process.

In the hypothesis testing problem, where we have to decide whether the target is present or absent, two kinds of errors can occur: 1) A false alarm which occurs when it is decided that the target is present when it is not, 2) A miss which occurs when it is decided that the target is not present when it is. In many radar problems the chosen criterion is to fix the probability of false alarm at a certain value and then to maximize the probability of detection. In statistical decision theory the *Likelihood Ratio Test (LRT)* is optimum for these kinds of problems. The *LRT* evaluates the likelihood ratio which is the ratio of the N^{th} order joint PDF under the alternative hypothesis H_1 (signal present) to the N^{th} order joint PDF under the null hypothesis H_0 (signal not present). This ratio is then compared to a certain threshold to make a decision. Under the constraint of a fixed false alarm probability, the Neyman-Pearson receiver obtained on the basis of the likelihood ratio test is the optimum receiver.

The components of the received vector $\underline{r}$ can be written mathematically as

$$H_1 : r_i = s_i + d_i \tag{1.1}$$

$$H_0 : r_i = d_i \quad i = 1, 2...N \tag{1.2}$$

where s_i and d_i represent the i^{th} sample of the desired signal return and the additive disturbance, respectively. Also, let $f_{\underline{R}}(\underline{r}|H_1)$, $f_{\underline{R}}(\underline{r}|H_0)$, $f_{\underline{D}}(\underline{d})$ denote the N^{th} order PDFs of $\underline{R}$ under H_1, of $\underline{R}$ under H_0 and of the disturbance. In general, the disturbance may be composed of clutter plus noise. Since it is not possible to separate the clutter and noise components of the disturbance when the disturbance is measured, we focus on the disturbance itself. As the signal becomes very weak (i.e. as the signal to clutter plus noise ratio (SCNR) approaches zero), the numerator and the

denominator of the LRT tend to become identical. This is due to the fact that

$$f_{\underline{R}}(\underline{r}|H_1) \approx f_{\underline{R}}(\underline{r}|H_0) = f_{\underline{D}}(\underline{d}). \tag{1.3}$$

This will result in the likelihood ratio being approximately equal to unity independent of the received signal. Thus, if T_s denotes the likelihood ratio,

$$P_D = \int_{\eta}^{\infty} f_{T_s}(T_s|H_1)dt_s \approx P_F = \int_{\eta}^{\infty} f_{T_s}(T_s|H_0)dt_s \tag{1.4}$$

where P_D and P_F represent the detection and false alarm probabilities. Therefore, the LRT performs poorly in the limit as the signal strength tends to zero.

Even though the problem of weak signal detection in radar applications is of great interest, most of the literature by various researchers has been devoted to strong signals in a clutter plus noise background. Optimal and/or very good sub-optimal schemes have been proposed to achieve the desired level of performance. Only a relatively small fraction of the literature is devoted to the design of practical schemes for the detection of weak signals. In this study we present a general theory for developing practical detector structures for weak signal problems. Also, computer simulations are used to evaluate performance when the disturbance can be approximated by the multivariate student-T and K-distributions. In such problems the concept of the Locally Optimum Detector (LOD) is used to come up with the decision rule which is also a ratio test. For a deterministic signal, a statistic is obtained by taking the ratio of the derivative with respect to the signal strength of the N^{th} order joint PDF under H_1 to the N^{th} order joint PDF under H_0. The limit of this ratio as the signal strength tends to zero is evaluated to obtain the test statistic for the decision rule. In the random signal case the test statistic is a ratio, in the limit as the signal strength tends to zero, of the second derivative with respect to the signal strength of the N^{th} order joint PDF under H_1 to the N^{th} order joint PDF under H_0. This approach is valid when it is known that the SCNR ratio is very small but the actual value of SCNR is unknown. Thus, the LOD turns out to be a Uniformly Most Powerful (UMP) test for the class of problems where the SCNR is in the neighborhood of zero.

1.2 Non-Gaussian Correlated Data

Previously, general analytic expressions for the various applicable N^{th} order joint non-Gaussian PDFs which allow for correlation between the variables were unavailable. As a result, researchers in the past assumed independence between the samples. By assuming independence between the samples, they were able to get the N^{th} order PDF as a product of the marginal density functions. If we carry out the locally optimum test using the N^{th} order density function based upon independence and evaluate its performance, it is found that an unreasonably large number of samples is needed for acceptable performance. This arises because independent samples imply a white spectrum. Consequently, space-time processing cannot be used to filter the target from the clutter. Based on the concept of Spherically Invariant Random Processes (SIRP), analytical expressions for some N^{th} order joint non-Gaussian PDFs which allow for correlation between the variables are now available. Use of the multivariate expressions for the non-Gaussian PDFs and the theory of locally optimum detectors enables receiver structures for weak signal detection to be derived.

1.3 Thesis Organization

The literature review on weak signal detection and the derivation of the locally optimum detector are presented in Chapter 2. It is shown that the LOD determines whether a target is present or absent by comparing a statistic computed from the data to a set threshold. Both deterministic and random target signals are considered. The receiver structures are specialized to the case for which the clutter plus noise can be approximated by an SIRP.

Since the clutter is assumed to be non-Gaussian, the LOD receiver structure turns out to be nonlinear. As a result, system performance must be determined by means of computer simulation. The threshold is conventionally determined through a Monte Carlo procedure. Unfortunately, the number of trials is inversely proportional to the false alarm probability, P_F. For example, when $P_F = 10^{-6}$, a minimum of ten million trials need to be generated. To avoid carrying out so many trials, a new technique, based on extreme value theory, is presented in Chapter 3. It is demonstrated that fairly accurate thresholds can be determined for false alarm probabilities as small as 10^{-7} with as few as 10,000-30,000 trials.

Assuming that the clutter plus noise can be approximated by either the multivari-

ate student-T or K-distributions, the LOD is developed in Chapter 4 for the weak signal detection problem. The system performance is evaluated by means of computer simulation for each distribution. It is shown that the performance improvement for the LOD is significant compared to the linear receiver when the clutter plus noise is approximated by the student-T distribution. However, the performance improvement compared to the linear receiver is not quite as significant when the clutter plus noise is approximated by the K-distribution.

To enhance the performance of the LOD for weak signal detection when the clutter plus noise is approximated as multivariate K-distributed, a new technique called the amplitude dependent locally optimum detector is presented. This test however, is not a uniformly most powerful test. Based on this test, it is demonstrated that significant performance improvements can be obtained compared to the linear receiver even when the clutter plus noise is multivariate K-distributed.

Summary and conclusions are presented in Chapter 6.

Chapter 2

The Locally Optimum Detector

2.1 Literature Review

The concept of the locally optimum detector was first established by Neyman and Pearson in 1930 [1, 2]. Subsequently this was applied to statistical communication and signal processing by several researchers.

David Middleton's work [3, 4] on the LOD is based on expanding the LRT in terms of a power series expansion and truncating the series to a first order approximation. In the limit as the signal tends to zero, the canonical structure of the locally optimum detector is established with very weak restrictions on the statistical properties of signal and noise. The analysis applies equally well to non-Gaussian as well as Gaussian and non-stationary as well as stationary processes, for stochastic as well as deterministic signals, continuous as well as discrete time signals and for combinations of signal and noise that need not be additive. In fact, the general character of the results is independent of the particular nature of the signal and noise, although specific noise distributions determine the specific detector structures. Middleton shows that the locally optimum detector is a threshold detector with very strong optimality features in the limit of an infinitely large number of samples. However, in our research, we are interested in applications where the number of samples may not be too large.

David Middleton [5] has also extended the problem of threshold or weak signal detection to vector fields involving highly non-Gaussian electromagnetic interference and signals which are both narrowband. The emphasis is on a canonical formulation. However, the performance measures presented are obtained on the basis of asymptotically locally optimum algorithms. In [6] Middleton has also analyzed the performance

of the locally optimum detectors in correlated interference. He points out that the correlation function involves only the second order statistics whereas higher order statistics should be considered for the non-Gaussian case. Consequently, the correlated noise model leads to a suboptimal receiver except when the underlying noise distribution is Gaussian. But when the sample to sample correlation is strong, the resulting algorithms and performance measures can provide noticeable improvements over models which employ independent noise sampling assumptions.

Other researchers in this area, such as J.H. Miller and John Thomas [7] and Saleem Kassam [8], have obtained performance of the LOD under the asymptotic condition of an infinitely large number of samples. These researchers have modeled the noise samples as independent, identically distributed random variables. This enables them to have a closed form expression for the N^{th} order PDF of multivariate non-Gaussian noise. Applying the LOD test they have arrived at the decision statistic. Using the central limit theorem, the test statistic is shown to approach Gaussian in the limit of very large sample size. Then the performance measures are evaluated. Song and Kassam [9], [10] have also derived the locally optimum detectors for both known signals and random signals in a generalized observation model. In this model additive, multiplicative and signal dependent noise models are considered. They show that the detectors derived under this model are interesting generalizations of the locally optimum detectors derived in the additive noise model case. They also analyze the performance of the detector for a finite sample size case. But, the underlying noise distribution is assumed to be bivariate Gaussian.

For a variety of detection problems, Jack Capon [11] concludes that implementation of the LOD is either less, or no more complicated than the Neyman-Pearson detector. First, he proposes the locally optimum detector for weak signal applications and proceeds to evaluate its performance by comparing it with the Neyman-Pearson detector. The comparison is based on the concept of Asymptotically Relative Efficiency (ARE). ARE is defined as the ratio of sample sizes required for two different detectors to achieve the same error probability and for the same signal to noise ratio, as the signal to noise ratio tends to zero and the sample sizes tend to infinity. On the basis of this comparison, it is shown that the locally optimum detector is asymptotically as efficient as the Neyman-Pearson detector. Conte, Izzo, Longo and Paura [12] have

also considered the problem of weak signal detection using locally optimum detectors for arbitrarily large sample sizes and show that significant improvements are achieved compared to the linear detector. However, when they implement their algorithm for finite sample sizes, they conclude that the promises of the asymptotic theory cannot be achieved using even moderately large sample sizes. Hence, they propose a scheme which is a hybrid of the asymptotically optimum detector and the linear detector. Asymptotically effective nonparametric algorithms for detecting weak signals in non-Gaussian interference have also been considered by Valeyev and Aspisov [13]. They conclude that the effectiveness of such algorithms approaches that of the optimum asymptotically, in the limit of large sample size. Raveendra and Srinivasan [14] have derived the locally optimum receiver structure for the coherent detection of continuous phase frequency shift keying (CPFSK) in non-Gaussian noise channels. They evaluate the performance of the receiver which consists of a zero memory nonlinearity followed by a correlator for a number of noise models. The measure of performance is the asymptotic relative efficiency. However, they point out two important drawbacks in the analysis of performance through the ARE. The first is that while large sample sizes are desirable for weak signal detection, increasing the sample size actually makes the LOD suboptimal partly due to the fact that there is an increased effect of higher order terms in the expansion of the likelihood ratio. Secondly, under increasing sampling rates, the assumption of independent samples, used in the derivation of the weak signal detector under non-Gaussian conditions becomes invalid. In a Naval Underwater Systems Center report, Raymond Ingram and R. Houle [15] analyze the performance of the optimum and several suboptimum receivers for weak signal detection of known signals in additive, white, non-Gaussian noise. He concludes that the implementation of the optimum or suboptimum nonlinear receivers yield significant improvements in performance relative to the receiver that is optimum in Gaussian noise. However, the receiver structure is more complicated than the linear receiver.

The structure of locally optimum detectors has also been characterized in terms of locally optimum estimators and correlators [16]. These characterizations are canonical structures involving estimators-correlators. It is shown that if the one step signal predictor is recursive and the noise is white (possibly non-Gaussian or nonstationary), there is a canonical structure that admits recursive computation. The motivation to

get a recursive structure is to simplify implementations and enhance adaptability. The problem of signal design has been considered for the case of locally optimum detection by Johnson and Orsak [17]. They show for the weak signal problem as we have defined it in Chapter 1 (Introduction), that the detection performance depends on signal energy in proportion to the Fisher information for location. In other words, when the spectra of the signal and disturbance overlap completely, significant improvements do not result from signal design. Another useful result they point out is that, among all distributions having zero mean and the same (finite) variance, the distribution having the smallest Fisher information is the Gaussian. Because of this result, it is concluded that detecting a small signal in Gaussian noise is the most difficult situation possible for an optimal detector. An increase in the signal energy yields the smallest possible performance improvement. Johnson and Orsak also come with explicit expressions to quantify the "small signal regimes" depending on the amplitude distribution of the noise.

Arthur Spaulding [18] compares the performance of the Locally Optimum Binary Detector (LOBD) with that of the linear receiver and a hard limiter. The performance analysis is done via specific examples and through computer Monte Carlo simulation. Under the assumption of independent, identically distributed samples, he concludes that the LOBD approaches its optimum performance only under the limit of large sample sizes and small signal to noise ratios. Also, he shows by way of an example that one cannot always be assured of obtaining great improvements over the linear receiver by using nonlinear processing. This implies that even if the underlying probability density function of the interference has a much larger tail than that of the Gaussian, it does not guarantee a much improved performance over the linear processor. The improvement in performance depends on the particular nature of the underlying PDF. Michael Bouvet [19] obtains the locally optimum detector by expanding the likelihood ratio and truncating the expansion under the weak signal assumption. He expands the likelihood ratio in two ways: one with respect to the signal and the other with respect to the observation. He then establishes the equivalence between the two different forms of expansion. However, he points out the limitations of these expansions by stating that the results are valid only if the neglected terms are actually negligible with respect to the retained terms. Since the received observations are random, this

cannot always be guaranteed.

Shishkov and Penev [20] have considered the case of correlated interference and background white noise, but have restricted themselves to multivariate Gaussian interference. For the known signal case, when the underlying noise distribution is Gaussian, the optimal detector obtained from the *LRT* is the same as the weak signal detector obtained from the locally optimum test. Modestino and Ningo [21] were among the earliest researchers to consider weak signal detection arising from bandpass processes. They have modeled the received signal as statistically independent complex samples and then obtained the joint density function of the inphase and quadrature components. Under the assumption that the clutter density function is circularly symmetric, they transform the joint density function to an equivalent one involving the envelope and phase. This model still does not include the correlation from one envelope sample to another and hence large sample sizes are required for good performance. Martinez, Swaszek and Thomas[22], have considered the case where the noise has a multivariate Laplace distribution, where any non-negative definite matrix can be used to model the correlation between the random variables. Based on this model they go ahead and derive the locally optimum detector which as expected is nonlinear. They compare the performance of this detector to the one developed by assuming the noise to be independent and identically distributed, and to the matched filter. They compare the performance of these receivers through the means of *ARE* and do not analyze the receiver performance for small sample sizes which is the case of practical interest.

Sangston and Gerlach [23] have used the the concept of the Spherically Invariant Random Processes to model multivariate non-Gaussian probability density functions. They derive the locally optimum detector based on this noise model. It turns out that the locally optimum detector structure is the matched filter in conjunction with a nonlinearity. They are able to establish the canonical nature of this result for the class of joint density functions arising from SIRPs. They propose an equivalent structure for the LOD. This takes the form of a receiver in which the matched filter output is compared to a nonlinear adaptive threshold. However, they do not explore the issue in terms of performance analysis.

2.2 Spherically Invariant Random Processes (SIRP)

In general, the radar receiver processes N complex samples (or $2N$ quadrature components) from each resolution cell. To develop an optimal receiver, it is necessary to have a closed form analytical expression for the joint PDF of the received samples. When the N samples are statistically independent, the joint PDF is simply the product of the marginal PDFs. However, clutter samples are likely to be correlated. Because this correlation is useful in canceling the clutter, it is important that the correlation be modeled. Unfortunately, when the received samples are correlated and non-Gaussian, there are no unique analytical expressions for their joint PDFs. A search of the mathematical and signal processing literature reveals that the SIRP provides a powerful mechanism for obtaining the PDF of N correlated non-Gaussian random variables. SIRPs for clutter modeling and simulation can be found in [24], where Conte and Longo model complex clutter as a spherically invariant random process. They point out that the SIRP is ergodic only if the underlying clutter process is Gaussian. This means that time averages cannot be used to approximate ensemble averages. Conte, Longo and Lops [25] also propose specific computer simulation procedures for generating clutter realizations from an SIRP with desired correlation properties when the underlying distribution is multivariate Weibull or K-distributed. Rangaswamy, Weiner and Ozturk [26] have developed a library of multivariate correlated non-Gaussian PDFs for characterizing various clutter scenarios through the theory of SIRPs. A significant result in this paper is the proof that the multivariate SIRP PDF approximation problem can be reduced to an equivalent univariate PDF approximation problem. In [27], Rangaswamy, Weiner and Ozturk develop two canonical computer simulation procedures for the generation of any correlated non-Gaussian clutter that can be modeled as a spherically invariant random process.

Application of the theory of SIRPs to the problem of signal detection and estimation can be found in [23] and [28]. In [28] Yao derives the form of the unit threshold likelihood ratio receiver for the detection of a known deterministic signal in additive SIRP noise. He shows that the optimum receiver is the linear receiver or the matched filter when the threshold is set to unity, a threshold that commonly arises in communication systems. This result is very significant because it tells us that nonlinear processing will not improve performance when the threshold is set to unity

even though the disturbance is non-Gaussian. However, when the threshold is not unity, then the optimal non-Gaussian receiver is a nonlinear receiver. Pentini, Farina and Zirilli [29] consider the problem of detecting a known target and a Swerling zero target embedded in coherent K-distributed clutter. The detectors are derived based on the likelihood ratio test where the multivariate joint density functions used in the test are obtained from the theory of SIRPs. The receiver performance is then evaluated for the strong signal case. The false alarm probabilities used in obtaining the receiver performance are 10^{-3} and 10^{-4} so as to reduce the number of Monte Carlo trials needed to set thresholds. For a Swerling zero target model, they obtain a probability of detection equal to 0.1 for a false alarm probability of 10^{-3} using four complex samples and a signal to clutter ratio of -10 dB. However, they do not explore performance for lower values of signal to clutter ratios.

2.3 The Derivation of the Locally Optimum Detector

The usual criterion in radar problems is to maximize the probability of detection under a fixed false alarm probability constraint. This receiver is called the Neyman-Pearson receiver. The receiver implements the Likelihood Ratio Test (LRT) and compares it against a threshold whose value is designed to give the desired false alarm probability. In particular, consider the received vector $\underline{R}^T = [R_1, R_2, ..., R_N]$. Introduce the two hypotheses H_0 and H_1 as described below:

$$H_0 : r_i = c_i + n_i \tag{2.1}$$

$$H_0 : r_i = \theta s_i + c_i + n_i \quad i = 1, 2...N. \tag{2.2}$$

Thus, H_0 is the hypothesis that the received signal consists solely of clutter plus noise while target signal is assumed to be present under the hypothesis H_1. Let the joint probability density function of $R_1, R_2, ..., R_N$ under hypothesis H_k $(k = 0, 1)$ be denoted by $f_{\underline{R}}(\underline{r}|H_k)$. The Neyman-Pearson receiver performs the LRT

$$t_s(\underline{r}) = \frac{f_{\underline{R}}(\underline{r}|H_1)}{f_{\underline{R}}(\underline{r}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.3}$$

where η is specified to satisfy the false alarm constraint

$$P_F = \int_{\eta}^{\infty} f_{T_s}(t_s|H_0) dt_s \tag{2.4}$$

and $f_{T_s}(t_s|H_k)$ is the conditional probability density function of the test statistic T_s given hypothesis H_k.

However, when the signal strength is very small relative to the clutter plus noise, the joint density function of the received random variables under H_1 approaches that under H_0. Then the numerator and the denominator of the LRT become approximately equal leading to numerical difficulties in discriminating between the two hypotheses. The Neyman-Pearson test is of course optimum. However, the form of the LRT can be rearranged to yield a test statistic which is more sensitive to perturbations in the received data. This gives rise to the concept of the Locally Optimum Detector (LOD). In this chapter the concept of the LOD is developed in detail using two approaches. The first approach is based on a power series expansion of the LRT and the second approach derives the LOD by an optimization using the principle of Lagrangian multipliers. It is shown that both approaches yield identical detector structures, though starting from different theoretical points of view. As the signal strength becomes weaker, the LOD becomes optimum even though its performance for a fixed sample size may not be as good as desired.

2.4 The Series Approach

2.4.1 The Known Signal Case

Let the additive clutter component $\underline{C} = [C_1, C_2, ..., C_N]^T$ be stationary and independent of the stationary white Gaussian background noise $\underline{N} = [N_1, N_2, ..., N_N]^T$. The noise variance σ_n^2 is assumed to be several orders of magnitude below the clutter variance σ_c^2 which is taken to be unity without loss of generality. The signal is assumed to be of the form $\theta \underline{s}$, where $\underline{s}$ is known. The components of $\underline{s}$ are chosen to have $|s_i|^2 = 1$ so that the positive parameter θ is a measure of the signal to clutter ratio (SCR) defined by

$$SCR = \frac{\theta^2 |s_i|^2}{\sigma_c^2} = \theta^2. \tag{2.5}$$

Because the clutter and noise are statistically independent with the noise assumed to have zero mean, the covariance matrix of the disturbance vector $\underline{D} = \underline{C} + \underline{N}$, denoted by M_D, is equal to the covariance matrix of the clutter M_C plus the covariance matrix of the noise M_N. Since the noise is white and stationary, the covariance matrix of the noise is of the form $M_N = \sigma_n^2 I$, where I is the identity matrix. When the clutter is highly correlated, the covariance matrix M_C tends to be ill-conditioned. However,

M_D will not be ill-conditioned because, by adding the small value σ_n^2 to the diagonal elements of M_C, the smallest eigenvalue of M_D is guaranteed to be no smaller than σ_n^2. Also, addition of M_N to M_C ensures that the disturbance spectrum will limit performance even in those frequency intervals where the clutter spectrum is negligible.

With this assumption the LRT takes the form

$$t_s = \frac{f_{\underline{R}}(\underline{r}|H_1)}{f_{\underline{R}}(\underline{r}|H_0)} = \frac{f_{\underline{D}}(\underline{r} - \theta\underline{s})}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{2.6}$$

As mentioned previously, when $\theta \ll 1$, the signal $\theta\underline{S}$ represents a small perturbation in the received vector under hypothesis H_1. Hence, $f_{\underline{R}}(\underline{r}|H_1)$ approximately equals $f_{\underline{R}}(\underline{r}|H_0)$. As a result, T_s is relatively insensitive to $\theta\underline{s}$. One approach at deriving a weak signal detector is to expand the numerator of the LRT in a Taylor series.

For this purpose, let $\underline{y} = \underline{r} - \theta\underline{s}$. Then

$$f_{\underline{R}}(\underline{r}|H_1) = f_{\underline{D}}(\underline{y}). \tag{2.7}$$

Expanding $f_{\underline{D}}(\underline{y})$ in a Taylor series about the received vector $\underline{r}$, we obtain

$$\begin{aligned} f_{\underline{D}}(\underline{y}) &= f_{\underline{D}}(\underline{r}) + \sum_{k_1=1}^{N} (y_{k_1} - r_{k_1}) \frac{\partial f_{\underline{D}}(\underline{y})}{\partial y_{k_1}}|_{\underline{y}=\underline{r}} \\ &+ \frac{1}{2!} \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} (y_{k_1} - r_{k_1})(y_{k_2} - r_{k_2}) \frac{\partial^2 f_{\underline{D}}(\underline{y})}{\partial y_{k_1} \partial y_{k_2}}|_{\underline{y}=\underline{r}} \\ &+ \cdots \\ &+ \frac{1}{n!} \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} \cdots \sum_{k_n=1}^{N} (y_{k_1} - r_{k_1})(y_{k_2} - r_{k_2}) \cdots (y_{k_n} - r_{k_n}) \frac{\partial^n f_{\underline{D}}(\underline{y})}{\partial y_{k_1} \partial y_{k_2} \cdots \partial y_{k_n}}|_{\underline{y}=\underline{r}} \\ &+ \cdots . \end{aligned} \tag{2.8}$$

This can be expressed in vector form by introducing the operator

$$(\underline{y} - \underline{r})^T \nabla_y = \sum_{k=1}^{N} (y_k - r_k) \frac{\partial}{\partial y_k} \tag{2.9}$$

where the subscript y on ∇ indicates partial differentiation with respect to the components of $\underline{y}$. The expansion of $f_{\underline{D}}(\underline{y})$ about the point $\underline{y} = \underline{r}$ then becomes

$$f_{\underline{D}}(\underline{y}) = f_{\underline{D}}(\underline{r}) + [(\underline{y} - \underline{r})^T \nabla_y] f_{\underline{D}}(\underline{y})|_{\underline{y}=\underline{r}}$$

$$
\begin{aligned}
&+ \frac{1}{2!}[(\underline{y}-\underline{r})^T\nabla_y]^2 f_{\underline{D}}(\underline{y})|_{\underline{y}=\underline{r}} \\
&+ \ldots \\
&+ \frac{1}{n!}[(\underline{y}-\underline{r})^T\nabla_y]^n f_{\underline{D}}(\underline{y})|_{\underline{y}=\underline{r}} \\
&+ \ldots \\
&= f_{\underline{D}}(\underline{r}) + \sum_{n=1}^{\infty}\frac{1}{n!}[(\underline{y}-\underline{r})^T\nabla_y]^n f_{\underline{D}}(\underline{y})|_{\underline{y}=\underline{r}}. \qquad (2.10)
\end{aligned}
$$

Recall that $\underline{y} = \underline{r} - \theta\underline{s}$, where θ and $\underline{s}$ are constants. Note that $\underline{y} - \underline{r} = -\theta\underline{s}$ and $\frac{\partial}{\partial y_k} = \frac{\partial}{\partial r_k}$. Then

$$(\underline{y}-\underline{r})^T\nabla_y = \sum_{k=1}^{N}(-\theta s_k)\frac{\partial}{\partial r_k} = -\theta\underline{s}^T\nabla_r \qquad (2.11)$$

where the subscript r on ∇ indicates partial differentiation with respect to the components of $\underline{r}$. It follows that the expansion may be written as

$$f_{\underline{D}}(\underline{r}-\theta\underline{s}) = f_{\underline{D}}(\underline{r}) + \sum_{n=1}^{\infty}\frac{(-1^n)}{n!}\theta^n[\underline{s}^T\nabla_r]^n f_{\underline{D}}(\underline{r}). \qquad (2.12)$$

In order for the above expansion to be meaningful, it is necessary that all the derivatives in the above expansion exist.

Thus, using the above expansion of $f_{\underline{D}}(\underline{r}-\theta\underline{s})$, the Taylor series expansion of the likelihood ratio about the received vector $\underline{r}$ in equation (2.6) can be written as

$$T_s(\underline{r}) = 1 + [\sum_{n=1}^{\infty}\frac{(-1)^n\theta^n}{n!\, f_{\underline{D}}(\underline{r})}(s^T\nabla_r)^n]f_{\underline{D}}(\underline{r}). \qquad (2.13)$$

The first term, being a constant, can be combined with the threshold without loss of optimality. The LOD is obtained by retaining only the term corresponding to $n = 1$ in the infinite summation. For $\theta \ll 1$, it is assumed that the remaining terms in the summation are negligible. On the other hand, because $\underline{r}$ is governed by a random vector and the partial derivatives of the PDF evaluated at $\underline{r}$ may be large, the remaining terms may actually not be negligible. However, it is assumed that this occurs with small probability. The resulting detector structure can then be expressed as

$$T_{LOD}(\underline{r}) = -\frac{(s^T\nabla_r)f_{\underline{D}}(\underline{r})}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_k \qquad (2.14)$$

where η_k is chosen so as to achieve the desired false alarm probability.

2.4.2 The Random Signal Case

When the signal is random, $f_{\underline{R}}(\underline{r}|H_1)$ is obtained by integrating the joint density function $f_{\underline{R},\underline{S}}(\underline{r},\underline{s}|H_1)$ over all possible values of $\underline{s}$. Hence,

$$f_{\underline{R}}(\underline{r}|H_1) = \int_{-\infty}^{\infty} f_{\underline{R},\underline{S}}(\underline{r},\underline{s}|H_1)d\underline{s} = \int_{-\infty}^{\infty} f_{\underline{R}|\underline{S}=\underline{s}}(\underline{r}|\underline{s},H_1)f_{\underline{S}}(\underline{s})d\underline{s} = E_s[f_{\underline{R}|\underline{S}=\underline{s}}(\underline{r}|\underline{s},H_1)] \tag{2.15}$$

where E_s denotes the expectation operation carried out with respect to the random vector $\underline{S}$. Because the denominator of T_s in equation (2.6) is independent of $\underline{s}$, the Taylor series expansion of the likelihood ratio can now be written as

$$T_s(\underline{r}) = 1 + [\sum_{n=1}^{\infty} \frac{(-1)^n\theta^n}{n!\, f_{\underline{D}}(\underline{r})} E_s[(\underline{s}^T\nabla_r)^n]f_{\underline{D}}(\underline{r})]. \tag{2.16}$$

Once again, as in the known signal case, the unity term appearing in the test statistic can be put into the threshold. If we make the assumption that the expected value of the signal vector is $\underline{0}$, then the $n = 1$ term in the infinite series of equation (2.16) goes to zero. Thus, for the random signal case, where the signal vector has zero mean, the LOD is defined to be the second term ($n = 2$) in the infinite series. As in the deterministic signal case, θ is assumed to be small enough such that the remaining terms of the series are negligible with high probability. Consequently, the LOD for the random signal case is given by

$$T_{s2}(\underline{r}) = \frac{\theta^2}{2f_{\underline{D}}(\underline{r})} E_s[(\underline{s}^T\nabla_r)^2]f_{\underline{D}}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta' \tag{2.17}$$

where T_{s2} represents the second order term in the Taylor series expansion of T_s. The above equation can be rewritten as

$$T_{s2}(\underline{r}) = \frac{\theta^2}{2f_{\underline{D}}(\underline{r})} E_s[\nabla_r^T\,\underline{s}\,\underline{s}^T\nabla_r]f_{\underline{D}}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta' \tag{2.18}$$

where, as before, η' is chosen to achieve the specified false alarm probability. Lumping the constant $\frac{\theta^2}{2}$ with the threshold and recognizing that

$$E_s[(\underline{s}^T\nabla_r)^2] = E_s[\nabla_r^T\,\underline{s}\,\underline{s}^T\nabla_r] = \nabla_r^T P\nabla_r, \tag{2.19}$$

where P is the covariance matrix of the signal vector, then the detector structure for the locally optimal test becomes

$$T_{LOD}(\underline{r}) = \frac{\nabla_r^T P \nabla_r [f_{\underline{D}}(\underline{r})]}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_u. \tag{2.20}$$

2.5 The Lagrangian approach

Consider again the hypotheses testing problem defined in equations (2.1-2.2). Let us define a nonrandomized decision rule $\phi(\underline{r})$ such that

$$\phi(\underline{r}) = \begin{cases} 1; & H_1 \textit{ true } (\textit{target present}) \\ 0; & H_0 \textit{ true } (\textit{target absent}). \end{cases} \tag{2.21}$$

This amounts to partioning the decision space into two regions, S_1 and S_0. A target is declared if the vector $\underline{r}$ is present in the region S_1. If it falls in the region S_0, then the decision is made that the target is absent. The probability of detection equals the probability that the nonrandomized decision rule equals unity, given that hypothesis H_1 is indeed true. This probability will, in general, be a function of θ, the signal-to-clutter ratio. Denoting $\beta(\theta)$ as the probability of detection we have

$$P_D = \beta(\theta) = p[\phi(\underline{r}) = 1|H_1] = \int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r}|H_1) \underline{dr}. \tag{2.22}$$

$\beta(\theta)$ is also called the power function of the test. The false alarm probability is given by

$$P_F = p[\phi(\underline{r}) = 1|H_0] = \int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r}|H_0) \underline{dr} = P_F. \tag{2.23}$$

The optimization problem to be discussed in the next section imposes the constraint that the false alarm probability be equal to P_F. P_F is also defined to be the significance level of the test.

2.5.1 The Known Signal Case

As discussed earlier, in the limit as the signal strength tends to zero, the probability of detection becomes approximately equal to the probability of false alarm. Therefore, instead of maximizing the probability of detection, one approach is to maximize the slope of the power function ($\beta(\theta)$) curve at the point θ equal to zero. The function to

be maximized and the constraint are given in the following two equations. Maximize

$$\frac{\partial \beta(\theta)}{\partial \theta}|_{\theta=0} = [\frac{\partial}{\partial \theta}\int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r}|H_1)\underline{dr}]|_{\theta=0} \tag{2.24}$$

subject to the constraint

$$\int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r}|H_0)\underline{dr} = \alpha. \tag{2.25}$$

We also require that the test be uniformly most powerful (UMP) in the sense that $\phi(\underline{r})$ be independent of θ for small neighborhoods in the vicinity of $\theta = 0$. Notice that there is a derivative with respect to θ outside the integral in equation (2.24). If the function $f_{\underline{R}}(\underline{r}|H_1)$ is a well behaved function of θ such that its derivative exists at all points, the derivative can be moved inside the integral resulting in

$$\frac{\partial}{\partial \theta}\int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r}|H_1)\underline{dr} = \int_{-\infty}^{\infty} \frac{\partial \phi(\underline{r})}{\partial \theta} f_{\underline{R}}(\underline{r}|H_1)\underline{dr} + \int_{-\infty}^{\infty} \phi(\underline{r})\frac{\partial f_{\underline{R}}(\underline{r}|H_1)}{\partial \theta}\underline{dr}. \tag{2.26}$$

Because of the UMP requirement, $\frac{\partial \phi(\underline{r})}{\partial \theta} = 0$ and the first integral on the right side of equation (2.26) integrates to zero. It follows that

$$\frac{\partial}{\partial \theta}\int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r}|H_1)\underline{dr} = \int_{-\infty}^{\infty} \phi(\underline{r})\frac{\partial f_{\underline{R}}(\underline{r}|H_1)}{\partial \theta}\underline{dr}. \tag{2.27}$$

Given the function $\frac{\partial \beta(\theta)}{\partial \theta}|_{\theta=0}$ to be maximized along with the false alarm probability constraint, the functional form of the maximization problem using the Lagrange multiplier approach is

$$max\ [\int_{-\infty}^{\infty} \phi(\underline{r})\frac{\partial f_{\underline{R}}(\underline{r}|H_1)}{\partial \theta}\underline{dr}|_{\theta=0} + \eta_k[\alpha - \int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r}|H_0)\underline{dr}]] \tag{2.28}$$

where η_k is the Lagrange multiplier. Expression (2.28) can be rewritten as

$$max\ [\int_{-\infty}^{\infty} \phi(\underline{r})[\frac{\partial f_{\underline{R}}(\underline{r}|H_1)}{\partial \theta} - \eta_k f_{\underline{R}}(\underline{r}|H_0)]\underline{dr}|_{\theta=0}] + \eta_k\alpha. \tag{2.29}$$

To maximize the above integral, the decision regions should be chosen such that the integrand is always positive. In other words, the decision regions are chosen such that

$$\frac{\partial f_{\underline{R}}(\underline{r}|H_1)}{\partial \theta}|_{\theta=0} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_k f_{\underline{R}}(\underline{r}|H_0). \tag{2.30}$$

As was pointed out in the previous section, $f_{\underline{R}}(\underline{r}|H_1)$ is identical to $f_{\underline{D}}(\underline{r}-\theta\underline{s})$. Therefore, the decision rule becomes

$$\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_k f_{\underline{D}}(\underline{r}). \tag{2.31}$$

The locally optimum detector is defined to be that detector which implements the ratio test

$$\frac{\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0}}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_k. \tag{2.32}$$

The Lagrange multiplier η_k is chosen to satisfy the false alarm constraint. Note that

$$f_{\underline{D}}(\underline{r}-\theta\underline{s}) = f_{\underline{D}}(r_1-\theta s_1, r_2-\theta s_2, ..., r_N-\theta s_N). \tag{2.33}$$

As a result,

$$\begin{aligned}\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta} &= \frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial(r_1-\theta s_1)}\frac{\partial(r_1-\theta s_1)}{\partial\theta} + \frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial(r_2-\theta s_2)}\frac{\partial(r_2-\theta s_2)}{\partial\theta} \\ &+ \ ... + \frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial(r_N-\theta s_N)}\frac{\partial(r_N-\theta s_N)}{\partial\theta} \\ &= \sum_{k=1}^{N}\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial(r_k-\theta s_k)}(-s_k).\end{aligned} \tag{2.34}$$

Consequently,

$$\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0} = -\sum_{k=1}^{N}\frac{\partial f_{\underline{D}}(\underline{r})}{\partial r_k}s_k = -(\underline{s}^T\nabla_r)f_{\underline{D}}(\underline{r}). \tag{2.35}$$

Thus, the locally optimum detector can also be written as

$$T_{LOD}(\underline{r}) = -\frac{(\underline{s}^T\nabla_r)f_{\underline{D}}(\underline{r})}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_k. \tag{2.36}$$

It can be seen that this detector is identical to the one in equation (2.14) obtained through the series approach.

2.5.2 The Random Signal Case

Consider a random signal $\underline{S}$ and let its joint PDF be denoted by $f_{\underline{S}}(\underline{s})$. Also, without loss of generality, we can make the assumption that the signal vector has zero mean and that each component of the vector has unit variance. Given the signal

vector $\underline{S}$ the joint density function on the received vector under hypothesis H_1 is

$$f_{\underline{R}}(\underline{r}|\underline{s}, H_1) = f_{\underline{D}}(\underline{r} - \theta\underline{s}). \tag{2.37}$$

The power function for the locally optimum test was given in the previous section in equation (2.22). However, in the random signal case the unconditional density function $f_{\underline{R}}(\underline{r}|H_1)$ is obtained by integrating out the random vector $\underline{S}$ from the joint PDF $f_{\underline{R},\underline{S}}(\underline{r}, \underline{s}|H_1) = f_{\underline{R}}(\underline{r}|\underline{s}, H_1)f_{\underline{S}}(\underline{s})$. Use of equation (2.37) results in

$$\beta(\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(\underline{r})f_{\underline{D}}(\underline{r} - \theta\underline{s})f_{\underline{S}}(\underline{s})\underline{dr}\ \underline{ds}. \tag{2.38}$$

The false alarm constraint is once again given by

$$\int_{-\infty}^{\infty} \phi(\underline{r})f_{\underline{R}}(\underline{r}|H_0)\underline{dr} = \alpha. \tag{2.39}$$

As before, we wish to maximize $\frac{\partial\beta(\theta)}{\partial\theta}|_{\theta=0}$. If the function $f_{\underline{D}}(\underline{r} - \theta\underline{s})$ is a well behaved function such that its derivative exists at all points, then

$$\frac{\partial\beta(\theta)}{\partial\theta} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(\underline{r})\frac{\partial f_{\underline{D}}(\underline{r} - \theta\underline{s})}{\partial\theta}f_{\underline{S}}(\underline{s})\underline{dr}\ \underline{ds}. \tag{2.40}$$

It follows from equation (2.35) that

$$\frac{\partial\beta(\theta)}{\partial\theta}|_{\theta=0} = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(\underline{r})[\sum_{k=1}^{N}\frac{\partial f_{\underline{D}}(\underline{r})}{\partial r_k}s_k]f_{\underline{S}}(\underline{s})\underline{dr}\ \underline{ds}. \tag{2.41}$$

Because of the zero mean assumption

$$\int_{-\infty}^{\infty} s_k f_{\underline{S}}(\underline{s})\underline{ds} = 0. \tag{2.42}$$

We conclude that

$$\frac{\partial\beta(\theta)}{\partial\theta}|_{\theta=0} = 0 \tag{2.43}$$

independent of the choice of $\phi(\underline{r})$. Therefore, to achieve the maximum increase of the power function in the vicinity of the origin, we maximize $\frac{\partial^2\beta(\theta)}{\partial\theta^2}|_{\theta=0}$. As before, assuming that the integration and differentiation can be interchanged,

$$\frac{\partial^2\beta(\theta)}{\partial\theta^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(\underline{r})\frac{\partial^2 f_{\underline{D}}(\underline{r} - \theta\underline{s})}{\partial\theta^2}f_{\underline{S}}(\underline{s})\underline{dr}\ \underline{ds}. \tag{2.44}$$

However from equation (2.34)

$$\begin{aligned}\frac{\partial^2 f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta^2} &= \frac{\partial}{\partial\theta}\sum_{k=1}^{N}\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial(r_k-\theta s_k)}(-s_k)\\ &= \sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial^2 f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial(r_j-\theta s_j)\partial(r_k-\theta s_k)}\frac{\partial(r_j-\theta s_j)}{\partial\theta}(-s_k)\\ &= \sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial^2 f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial(r_j-\theta s_j)\partial(r_k-\theta s_k)}s_j s_k. \qquad (2.45)\end{aligned}$$

Hence,

$$\frac{\partial^2 f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta^2}|_{\theta=0} = \sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial^2 f_{\underline{D}}(\underline{r})}{\partial r_j\partial r_k}s_j s_k = (\nabla_r^T\underline{s}\,\underline{s}^T\nabla_r)f_{\underline{D}}(\underline{r}). \qquad (2.46)$$

Then the second derivative of the power function at the origin takes the form

$$\frac{\partial^2\beta(\theta)}{\partial\theta^2}|_{\theta=0} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(\underline{r})(\nabla_r^T\underline{s}\,\underline{s}^T\nabla_r)f_{\underline{D}}(\underline{r})f_{\underline{S}}(\underline{s})d\underline{r}\,d\underline{s} = \int_{-\infty}^{\infty}\phi(\underline{r})E_s[\nabla_r^T\underline{s}\,\underline{s}^T\nabla_r]f_{\underline{D}}(\underline{r})\,d\underline{r}. \qquad (2.47)$$

Using the approach of Lagrange multipliers to maximize the function in equation (2.47) along with the constraint (2.39), the optimization problem can be written as

$$max[\int_{-\infty}^{\infty}\phi(\underline{r})E_s[\nabla_r^T\underline{s}\,\underline{s}^T\nabla_r]f_{\underline{D}}(\underline{r})d\underline{r} + \eta_u[\alpha - \int_{-\infty}^{\infty}\phi(\underline{r})f_{\underline{D}}(\underline{r})d\underline{r}]]. \qquad (2.48)$$

The above expression can be rewritten as

$$max[\int_{-\infty}^{\infty}\phi(\underline{r})[E_s[\nabla_r^T\underline{s}\,\underline{s}^T\nabla_r]f_{\underline{D}}(\underline{r}) - \eta_u f_{\underline{D}}(\underline{r})]d\underline{r}] + \eta_u\alpha. \qquad (2.49)$$

To maximize the integral the decision regions have to be chosen such that the integrand is always nonnegative. The resulting decision regions yield the inequalities

$$E_s[\nabla_r^T\underline{s}\,\underline{s}^T\nabla_r]f_{\underline{D}}(\underline{r}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_u f_{\underline{D}}(\underline{r}). \qquad (2.50)$$

If the covariance matrix of the signal vector is denoted by P, then the locally optimum detector can be written as

$$T_{LOD}(\underline{r}) = \frac{(\nabla_r^T P\nabla_r)f_{\underline{D}}(\underline{r})}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_u \qquad (2.51)$$

which is identical to equation (2.20).

As a general rule for deriving locally optimum tests, note that we maximize at the

origin the first non-vanishing derivative of the power function. For the known and the purely random signal cases the first non-vanishing derivative is the first and the second derivative, respectively.

2.6 Special Cases

In this section LOD structures will be derived for three special cases. In the first it is assumed that the N random variables in the disturbance vector $\underline{D}$ are statistically independent. With this assumption, the joint PDF of the N random variables is the product of the marginal density functions of the individual random variables. In the second, the N random variables are modeled as arising from an SIRP. This model enables us to write the joint PDF of the random variables analytically, accounting for the correlation between the random variables. In the last case the N correlated random variables are assumed to be jointly Gaussian. The locally optimum detector structures are derived for all cases. It turns out in all three cases that the detector can be expressed in a canonical form. This canonical expression is derived for both the known and the random signal problems.

2.6.1 The Known Signal Problem

2.6.1.1 Disturbance Modeled as Independent Random Variables

From equation (2.32), the LOD structure in the known signal case is given as

$$\frac{\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0}}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_k. \tag{2.52}$$

Let the N random variables in the vector $\underline{D}$ be independent where the PDF of the i^{th} random variable is $f_{D_i}(d_i)$. Therefore, the conditional joint density functions of the N received random variables are given by

$$f_{R_1,R_2,....,R_N}(r_1,r_2,...,r_N|H_0) = \prod_{i=1}^{N} f_{D_i}(r_i) \tag{2.53}$$

$$f_{R_1,R_2,....,R_N}(r_1,r_2,...,r_N|H_1) = \prod_{i=1}^{N} f_{D_i}(r_i-\theta s_i). \tag{2.54}$$

The numerator in the ratio test of equation (2.52) is evaluated as

$$\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0} = \frac{\partial}{\partial\theta}[\prod_{i=1}^{N} f_{D_i}(r_i-\theta s_i)]|_{\theta=0} = \sum_{i=1}^{N}\{(-s_i)\frac{df_{D_i}(r_i)}{dr_i}\prod_{\substack{j=1\\ j\neq i}}^{N} f_{D_j}(r_j)\}. \tag{2.55}$$

Thus, from equation (2.52) the LOD statistic for independent random variables is given by

$$T_{LOD}(r_1, r_2, ..., r_N) = -\sum_{i=1}^{N} s_i \frac{f'_{D_i}(r_i)}{f_{D_i}(r_i)} \tag{2.56}$$

where $f'_{D_i}(r_i)$ denotes the derivative of $f_{D_i}(r_i)$ with respect to r_i. The above equation for the LOD statistic is the canonical form obtained when the random variables are independent. For different density functions, $f_{D_i}(r_i)$, the detector will be different, although its structure remains the same. The canonical form of the detector is shown in Fig. 2.1.

2.6.1.2 Disturbance Modeled as an SIRV

When the random variables of the disturbance are drawn from a zero mean SIRP distribution, the joint PDF can be written as

$$f_{\underline{D}}(\underline{d}) = \frac{1}{2\pi^{N/2}|M|^{1/2}} h_N(p) \tag{2.57}$$

where $p = \underline{d}^T M^{-1} \underline{d}$, M is the covariance matrix for the N random variables and $h_N(p)$ is a positive valued, nonlinear function of p. The numerator of the ratio test in equation (2.52) is then given by

$$\frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta}|_{\theta=0} = \frac{\partial}{\partial \theta}\{\frac{1}{2\pi^{N/2}|M|^{1/2}} h_N(p)\}|_{\theta=0} = \frac{1}{2\pi^{N/2}|M|^{1/2}} \frac{\partial}{\partial \theta}\{h_N(p)\}|_{\theta=0}. \tag{2.58}$$

where the quadratic form p equals $(\underline{r} - \theta \underline{s})^T M^{-1} (\underline{r} - \theta \underline{s})$ since $\underline{d} = \underline{r} - \theta \underline{s}$. From the chain rule for differentiation we have

$$\frac{\partial}{\partial \theta}(h_N(p)) = \frac{\partial}{\partial p}(h_N(p)) \frac{\partial p}{\partial \theta}. \tag{2.59}$$

From the expression for p

$$\frac{\partial p}{\partial \theta}|_{\theta=0} = -2(\underline{s}^T M^{-1} \underline{r}). \tag{2.60}$$

Making use of equations (2.58-2.60) the LOD statistic in equation (2.52) becomes

$$T_{LOD}(\underline{r}) = -2(\underline{s}^T M^{-1} \underline{r}) \frac{h'_N(p)}{h_N(p)} \tag{2.61}$$

where $h'_N(p)$ denotes the derivative of the function $h_N(p)$ with respect to the argument p. The LOD statistic in equation (2.61) represents the canonical structure when

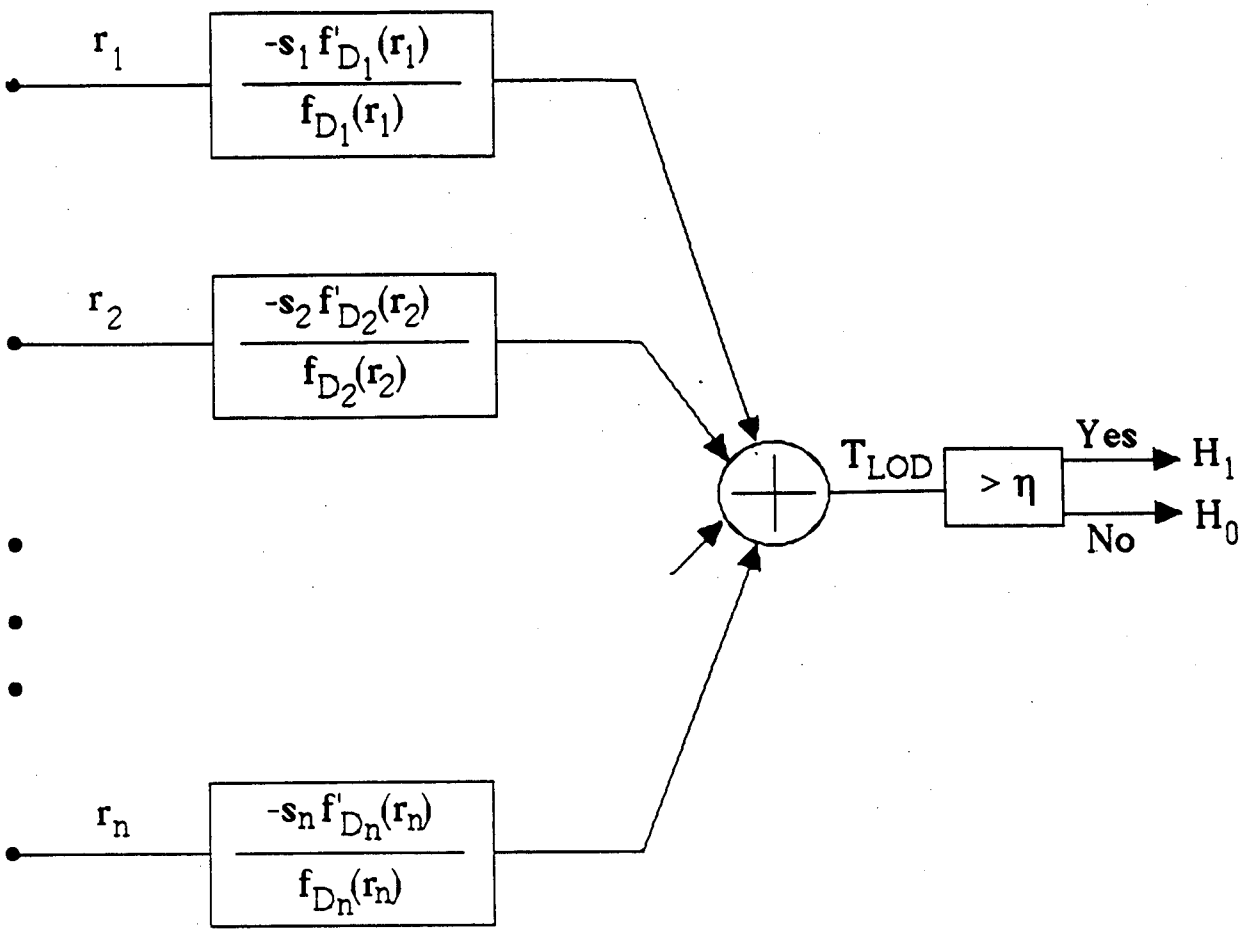


Figure 2.1: Canonical form of LOD assuming known signal and independent random variables.

the disturbance is modeled as an SIRV. The nonlinear function $h_N(p)$ depends on the particular joint density function used to model the disturbance. The canonical structure for the detector is shown in Fig. 2.2. Note that the detector multiplies the output of a matched filter with the output of a nonlinearity. Just as with a Gaussian receiver, the matched filter maximizes the signal to disturbance ratio even though the received signal is non-Gaussian (i.e., derivation of the matched filter to maximize signal to noise ratio does not depend on the Gaussian assumption). For non-Gaussian problems, matched filtering alone is suboptimum. For SIRPs the optimal receiver requires nonlinear processing as well as matched filtering.

2.6.1.3 Random Variables Arising from the Gaussian Distribution

The SIRP class of distributions reduces to the Gaussian distribution when

$$h_N(p) = e^{-\frac{p}{2}}. \tag{2.62}$$

It follows that

$$\frac{h_N'(p)}{h_N(p)} = -\frac{1}{2}. \tag{2.63}$$

With reference to equation (2.61), the LOD statistic becomes

$$T_{LOD}(\underline{r}) = \underline{s}^T M^{-1} \underline{r}. \tag{2.64}$$

Interestingly enough, this is identical to the statistic of the likelihood ratio test for the known signal Gaussian problem [30]. Hence, for the known signal Gaussian problem, the strong and the weak signal detectors are identical. Note that there is no nonlinearity involved with the weak signal detector for this case. To put it another way, the general nonlinear SIRP weak signal detector of Fig. 2.2 reduces to the linear receiver or matched filter known to be optimum for the Gaussian problem.

2.6.2 The Random Signal Problem

2.6.2.1 Independent Disturbance Random Variables

The locally optimum detector is given by equation (2.51) when the signal is random. Repeating equation (2.51) the LOD structure is

$$T_{LOD}(\underline{r}) = \frac{(\nabla_r^T P \nabla_r) f_{\underline{D}}(\underline{r})}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_u. \tag{2.65}$$

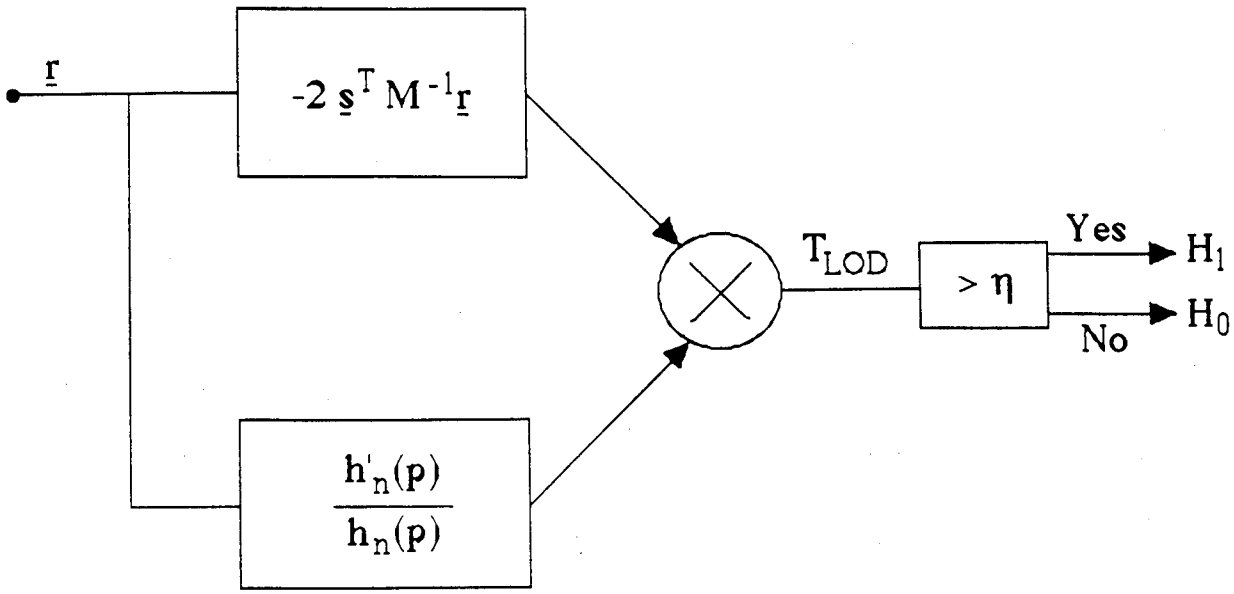


Figure 2.2: Canonical form of LOD assuming known signal and random variables arising from an SIRP.

P is the random signal covariance matrix. In this section the components of the disturbance vector $\underline{D}$ are assumed to be statistically independent. The analysis is further simplified when the signal random variables are also assumed to be uncorrelated. The covariance matrix P then becomes diagonal. Let the diagonal elements of the matrix P be represented by σ_i^2, $i = 1, 2, ..., N$. Because the disturbance random variables are independent, the joint density function $f_{\underline{D}}(\underline{r})$ is again given by the product of the marginal density functions of the individual random variables. Specifically,

$$f_{\underline{D}}(\underline{r}) = \prod_{i=1}^{N} f_{D_i}(r_i). \tag{2.66}$$

Also, when P is diagonal,

$$\nabla_r^T P \nabla_r = \sum_{i=1}^{N} \sigma_i^2 \frac{\partial^2}{\partial r_i^2}. \tag{2.67}$$

Using equations (2.65-2.67) and following the same steps as in the known signal case, the LOD statistic can be derived as

$$T_{LOD}(\underline{r}) = \sum_{i=1}^{N} \sigma_i^2 \frac{f''_{D_i}(r_i)}{f_{D_i}(r_i)} \tag{2.68}$$

where the double prime indicates second derivative with respect to the argument. The canonical structure derived above is shown in Fig. 2.3.

2.6.2.2 Disturbance Random Variables Arising from an SIRP Distribution

When the disturbance vector is modeled as having an SIRP distribution, the joint PDF is given by equation (2.57). The LOD structure for the random signal case is given by equation (2.65). Since the constant terms in the joint density function cancel out in the numerator and denominator of the ratio test in equation (2.65), the LOD statistic is obtained by evaluating

$$T_{LOD}(\underline{r}) = \frac{(\nabla_r^T P \nabla_r) h_N(p)}{h_N(p)}. \tag{2.69}$$

The numerator of equation (2.69) can be expanded as a sum of terms involving partial derivatives. The result is simplified considerably when the covariance matrix P of the signal vector is diagonal. When P is chosen to be the Identity matrix (i.e. P is diagonal and the variance of each element of the signal vector is unity), the LOD

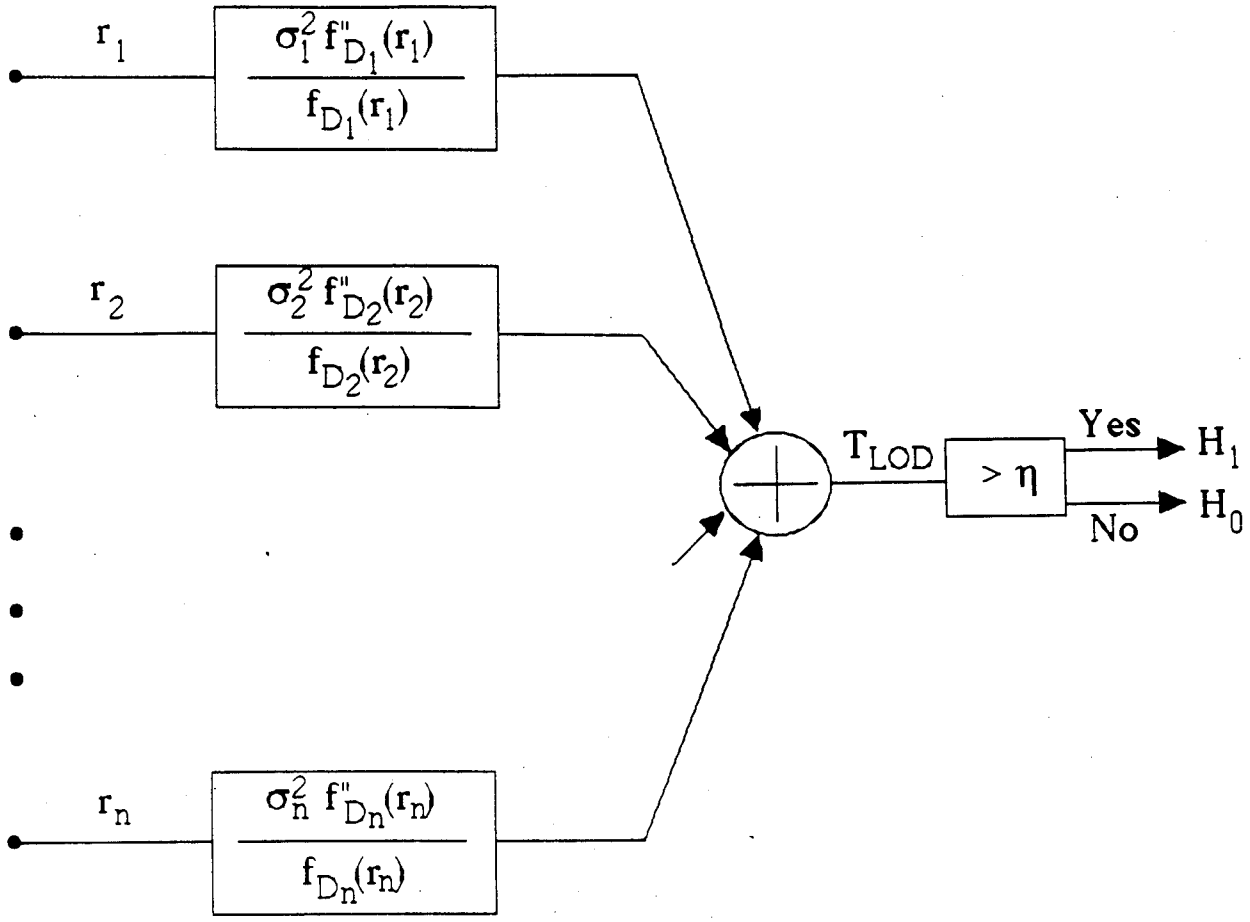


Figure 2.3: Canonical form of LOD assuming random signal and independent random Variables.

statistic is given by

$$T_{LOD}(\underline{r}) = \frac{(\nabla_r^T \nabla_r) h_N(p)}{h_N(p)}. \tag{2.70}$$

The inner product involving the ∇ vector can be written as

$$\nabla_r^T \nabla_r = \sum_{i=1}^{N} \frac{\partial^2}{\partial r_i^2}. \tag{2.71}$$

Application of equation (2.71), to the the numerator of equation (2.70) results in

$$\sum_{i=1}^{N} \frac{\partial^2 h_N(p)}{\partial r_i^2} = \sum_{i=1}^{N} h_N'(p) \frac{\partial^2 p}{\partial r_i^2} + h_N''(p) (\frac{\partial p}{\partial r_i})^2 \tag{2.72}$$

where the prime indicates differentiation with respect to p. Using equation (2.72) and dividing by $h_N(p)$ the LOD statistic becomes

$$T_{LOD}(\underline{r}) = \frac{1}{h_N(p)} [\sum_{i=1}^{N} (h_N'(p) \frac{\partial^2 p}{\partial r_i^2} + h_N''(p) (\frac{\partial p}{\partial r_i})^2). \tag{2.73}$$

The quadratic form p can be written as

$$p = \sum_{k=1}^{N} \sum_{l=1}^{N} r_k M_{kl}^{-1} r_l \tag{2.74}$$

where M_{kl}^{-1} represents the k^{th} row and l^{th} column entry of the matrix M^{-1}. From equation (2.74) $\frac{\partial p}{\partial r_i}$, $(\frac{\partial p}{\partial r_i})^2$ and $\frac{\partial^2 p}{\partial r_i^2}$ can be calculated. In particular, we have

$$\begin{aligned} \frac{\partial p}{\partial r_i} &= \frac{\partial}{\partial r_i} \sum_{k=1}^{N} \sum_{l=1}^{N} r_k M_{kl}^{-1} r_l \\ &= \sum_{k=1}^{N} r_k M_{ki}^{-1} + \sum_{l=1}^{N} r_l M_{il}^{-1}. \end{aligned} \tag{2.75}$$

Because of the symmetric nature of the matrices M and M^{-1}, $M_{ki}^{-1} = M_{ik}^{-1}$. It follows that the square of the above equation is then given by

$$\begin{aligned} (\frac{\partial p}{\partial r_i})^2 &= (\sum_{k=1}^{N} r_k M_{ki}^{-1} + \sum_{l=1}^{N} r_l M_{il}^{-1})^2 = (2 \sum_{k=1}^{N} r_k M_{ki}^{-1})^2 \\ &= 4 \sum_{k=1}^{N} \sum_{l=1}^{N} r_k M_{ki}^{-1} M_{li}^{-1} r_l = 4 \sum_{k=1}^{N} \sum_{l=1}^{N} r_k M_{ki}^{-1} M_{il}^{-1} r_l \end{aligned} \tag{2.76}$$

Utilizing equation (2.75),

$$\frac{\partial^2 p}{\partial r_i^2} = \frac{\partial}{\partial r_i}(\sum_{k=1}^{N} r_k M_{ki}^{-1} + \sum_{l=1}^{N} r_l M_{il}^{-1}) = 2M_{ii}^{-1}. \tag{2.77}$$

With reference to equations (2.73) and (2.75-2.77) define

$$T_{LOD}^{(1)}(\underline{r}) = \frac{1}{h_N(p)}[\sum_{i=1}^{N} h_N'(p)\frac{\partial^2 p}{\partial r_i^2}] = 2\frac{h_N'(p)}{h_N(p)}\sum_{i=1}^{N} M_{ii}^{-1}, \tag{2.78}$$

$$\begin{aligned} T_{LOD}^{(2)}(\underline{r}) &= \frac{1}{h_N(p)}\sum_{i=1}^{N} h_N''(p)(\frac{\partial p}{\partial r_i})^2 = 4\frac{h_N''(p)}{h_N(p)}\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} r_k M_{ki}^{-1} M_{il}^{-1} r_l \\ &= \frac{4h_N''(p)}{h_N(p)}\underline{r}^T M^{-1} M^{-1}\underline{r}. \end{aligned} \tag{2.79}$$

The locally optimum detector statistic that results from equations (2.78-2.79) is written as

$$\begin{aligned} T_{LOD}(\underline{r}) &= T_{LOD}^{(1)}(\underline{r}) + T_{LOD}^{(2)}(\underline{r}) \\ &= \frac{1}{h_N(p)}[2h_N'(p)tr(M^{-1}) + 4h_N''(p)\underline{r}^T M^{-1} M^{-1}\underline{r}] \end{aligned} \tag{2.80}$$

where $tr(M^{-1})$ is the sum of the all the diagonal elements of the matrix M^{-1}. The canonical structure of the receiver is shown in Fig. 2.4.

2.6.2.3 Disturbance Random Variables Arising from the Gaussian Distribution

As pointed out in section 2.6.1.3, the SIRP distribution reduces to the Gaussian distribution when

$$h_N(p) = e^{-\frac{p}{2}}. \tag{2.81}$$

From the above equation it follows that

$$\begin{aligned} \frac{h_N'(p)}{h_N(p)} &= -\frac{1}{2} \\ \frac{h_N''(p)}{h_N(p)} &= \frac{1}{4}. \end{aligned} \tag{2.82}$$

Consequently equation (2.78) reduces to

$$T_{LOD}^{(1)}(\underline{r}) = -tr(M^{-1}) \tag{2.83}$$

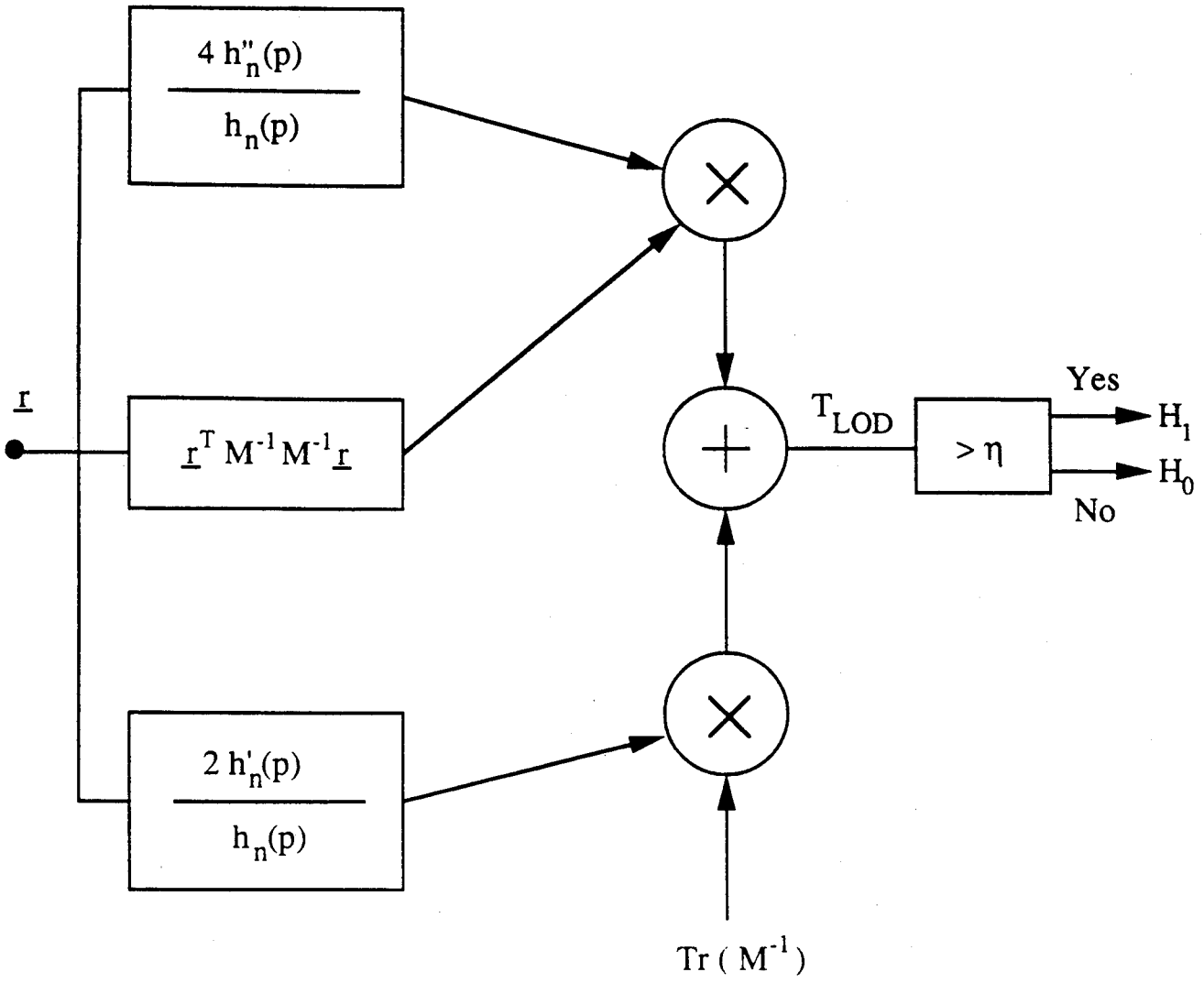


Figure 2.4: Canonical form of LOD assuming random signal and random disturbance arising from an SIRP.

whereas equation (2.79) becomes

$$T_{LOD}^{(2)}(\underline{r}) = \underline{r}^T M^{-1} M^{-1} \underline{r}. \tag{2.84}$$

Note that $T_{LOD}^{(1)}(\underline{r})$ is a constant that can be combined with the threshold. As a result, the LOD statistic for the random signal Gaussian problem is given by

$$T_{LOD}(\underline{r}) = \underline{r}^T M^{-1} M^{-1} \underline{r}. \tag{2.85}$$

Unlike the known signal Gaussian problem, the weak signal LOD statistic does not equal the statistic of the likelihood ratio for the random signal Gaussian problem [30] which is

$$T_{LR}(\underline{r}) = \underline{r}^T [M^{-1} - (M + P)^{-1}] \underline{r}. \tag{2.86}$$

The two statistics become equivalent only when

$$M = P = I. \tag{2.87}$$

Although the strong and weak signal detectors have different test statistics, the receiver structures are both quadratic in nature.

Chapter 3

Determining Thresholds for the Locally Optimum Detector

3.1 Literature Review

In this dissertation, multivariate density functions for modeling non-Gaussian probability density functions are assumed to be known. They are obtained using the theory of spherically invariant random processes. Once the multivariate density functions are known, we derive a decision rule using the theory of locally optimum detectors, that is applicable when the signal to be detected is weak compared to the additive disturbance. The procedure for obtaining the decision rule is explained in detail in Chapter 2.

However, the detector that is obtained on the basis of the theory of locally optimum detectors is typically nonlinear as the underlying processes are non-Gaussian. When the test statistic is nonlinear, it is not possible to evaluate the performance of the detector analytically. Consequently, we have to resort to computer simulations to analyze the performance. There are two steps involved in computer simulations to analyze performance. The first step is to evaluate the threshold so as to obtain the desired false alarm probability. The second step is to evaluate the detection probability once the threshold is set, corresponding to the desired false alarm probability.

3.1.1 Classical methods for evaluating thresholds

Monte Carlo methods have typically been used for this purpose. A large number of trials M are generated under the hypothesis that the received signal consists of the disturbance alone. The output of the detector, T_{s_i} $i = 1, 2, ..., M$ corresponding to

the generated disturbance vectors are recorded. Based on the output of the detector, thresholds can be set to obtain the desired false alarm probability. But, in order to establish the threshold for a specified P_F, it is necessary to accurately know the behavior of the tail of the test statistic. Unfortunately, the number of trials required for the Monte Carlo technique is very large, as is evident from the rule of thumb

$$M \geq \frac{10}{P_F}. \tag{3.1}$$

Hence, if $P_F = 10^{-5}$, one million trials should be generated. Clearly this is not a very desirable situation. Thus, for a reasonable sample size M, estimation of thresholds corresponding to small false alarm probabilities cannot be made when these methods are used. For Monte Carlo simulations the construction of approximate confidence intervals for the threshold estimates based on various estimators are discussed by Hosking and Wallis [31].

Some other approaches which do not make use of raw data or their smoothed versions have been suggested by various authors. For example, Harrel and Davis [32] suggested using linear combinations of sample order statistics. Their approach appears to be applicable for estimation of thresholds in the central region of the distribution. However, the underlying threshold estimator is not in a simple computational form and does not offer any additional advantage over the Monte Carlo method in terms of threshold estimation corresponding to small false alarm probabilities.

It has recently been shown [33] that the PDFs of the test statistic can be determined experimentally using a relatively small number of samples (eg: 50-100 samples give accurate fits depending on the distribution). Because the number of samples required by Ozturk's technique is small, it is unlikely that samples will be from the extreme tails of the PDFs. Consequently, the accurate fit mentioned above applies to the main body of the density function.

A number of statisticians have developed methods for estimating the extreme tail of the distributions using the asymptotic properties of extreme order statistics. Assuming an unknown probability density function $f_X(x)$, then for large X Hill [34] proposed using $\Lambda(x) = 1 - cx^{-\alpha}$ as a limiting form of the distribution function to infer the tail behavior. A similar approach is also given by Pickands [35]. Weissman [36] proposed a different approach based on the joint limiting distribution of the

largest k order statistics. His approach is based on the fact that "the largest k order statistics have a joint distribution that is approximately the same as the largest k order statistics from a location and scale exponential distribution ". Weissman obtained a simple expression for the estimate of the thresholds corresponding to various false alarm probabilities of the distributions. Based on the empirical comparisons, Boos [37] reported that Weissman's estimators have lower mean squared error than those of standard methods when the tails are exactly exponential. When the tails are not exactly exponential the estimators become highly biased. The mean squared errors of the estimators strongly depend on the choice for k. Although the method is non-parametric in nature, the optimal choice of k requires the knowledge of the parent distribution.

The use of stable distributions to model data having large tailed distributions has attracted considerable attention [38], [39],[40], and [41]. The independent and identically distributed random variables $Y_1, Y_2, ...Y_n$ are said to have a stable distribution if $Y_1 + Y_2 + ... + Y_n$ has the same distribution as the individual random variables. With the introduction of additional parameters, control of the mean, variance and the skewness of the distribution is possible. A major difficulty with stable distributions is that they usually cannot be expressed in closed form. Also, estimation of parameters is not computationally easy [42].

3.2 Extreme value theory

Guida et al. [43] compared the performance analysis of some extrapolative estimators of probability tails with application to radar systems design. They show that the estimates based on the extreme value theory yield clearly superior accuracy, while achieving a substantial savings in sample size compared to the classical Monte Carlo techniques. However, their method suffers from two major drawbacks. First, they assumed that the underlying unknown distribution is always exponential in nature. This assumption can be restrictive in certain situations. The other drawback is that the samples are partitioned into many smaller sets of samples and the maximum from each set is drawn for estimation purposes. They provide no optimum rule for determining the number of sets to be used in partitioning the original sample even though the accuracy of the estimation depends strongly on the original sample size and the number of sets.

Pickands [35] first suggested that the Generalized Pareto distribution can be used to model the extreme tails of probability density functions. The Generalized Pareto Distribution (GPD) is a two parameter distribution, with a scale and a shape parameter. Modeling the extreme tail then corresponds to estimating the two parameters of the GPD. The estimation methods for the Generalized Pareto Distribution have been reviewed by Hosking and Wallis [31]. They considered the method of moments, probability weighted moments and the maximum likelihood method for estimating the parameters and the thresholds. Based on computer simulation experiments, they showed that the probability weighted moment method is more reliable than the maximum likelihood method for sample sizes less than 500.

3.3 The Radar Problem

The hypothesis testing problem for deciding whether or not a target is present is given by equations (2.1-2.2) in Chapter 2. For weak signal applications, it was shown that the Locally Optimum Detector is useful for arriving at a decision rule. For the known signal case, the LOD structure is given by equation (2.32). Since the test statistic is typically a nonlinear function when $f_{\underline{D}}(\underline{r}|H_0)$ and $f_{\underline{D}}(\underline{r}|H_1)$ are multivariate non-Gaussian density functions, it is not possible, in general, to analytically evaluate in closed form the threshold η for a specified false alarm probability. Given the probability density functions (PDF) of the test statistic, T_s, under hypotheses H_1 and H_0, the detection and false alarm probabilities are

$$P_D = \int_{\eta}^{\infty} f_{T_s}(t_s|H_1)dt_s \tag{3.2}$$

$$P_F = \int_{\eta}^{\infty} f_{T_s}(t_s|H_0)dt_s. \tag{3.3}$$

P_D and P_F are represented by the shaded areas shown in Fig. 3.1. As indicated in the figure P_F is typically much smaller than P_D.

In practice, the density function of T_s is not known in advance. For example, depending upon various conditions such as terrain, weather etc., the clutter may be best modeled by Gaussian, K-distributed, Weibull or some other probability distribution. In this Chapter a new approach is developed for experimentally determining the extreme tail of $f_{T_s}(t_s|H_0)$, where the number of samples required is several orders of magnitude smaller than that suggested by equation (3.1). Once the tail of $f_{T_s}(t_s|H_0)$

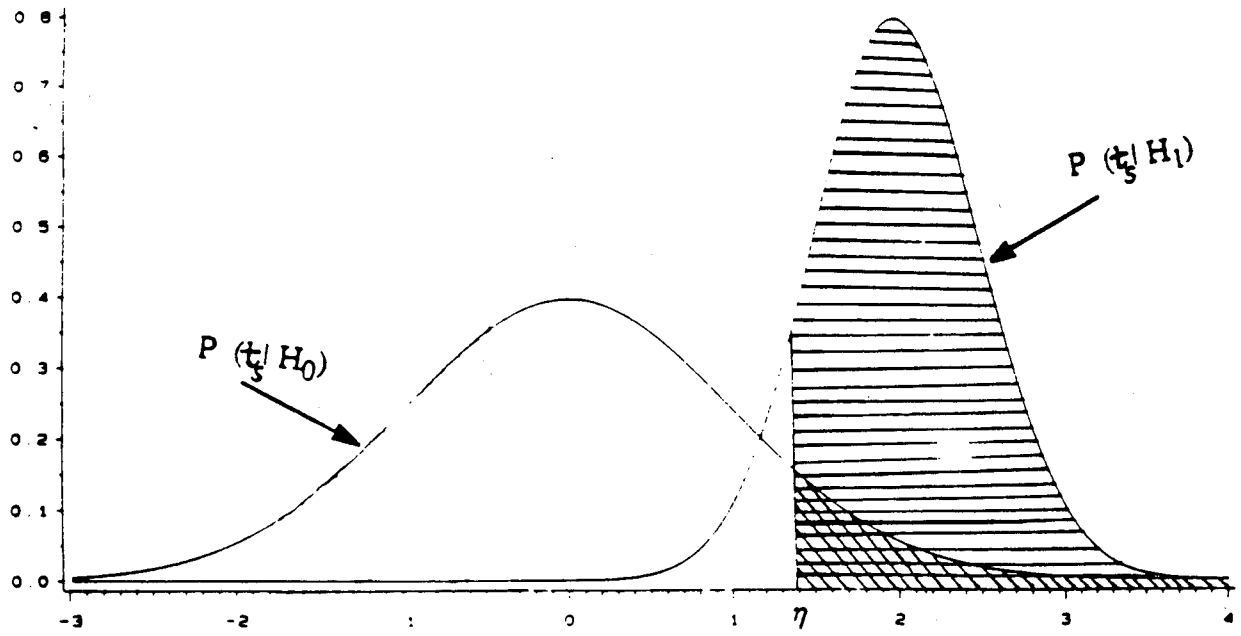


Figure 3.1: Shaded areas indicating P_F and P_D.

has been estimated, the threshold can be determined by use of equation (3.3).

3.4 Methods for Estimating Thresholds

3.4.1 Estimates Based on Raw Data

In this section we consider some commonly used threshold estimates. These estimates are called raw estimates and are already included in some statistical package programs (eg: the UNIVARIATE procedure in the SAS [44] package).

Let $X_1 \leq X_2 \leq ... \leq X_n$ denote the sample order statistics from a distribution function $F(x)$. Let p denote the desired false alarm probability. Also, let $n(1-p) = j + g$ where j is the integer part of $n(1-p)$. We denote the threshold estimate based on the k^{th} procedure to be described below by $\eta_p^{(k)}$. Four different threshold estimates are given as follows:

$$\eta_\alpha^{(1)} = (1-g)X_j + gX_{j+1} \tag{3.4}$$

$$\eta_\alpha^{(2)} = X_k, \textit{ where } k \textit{ is the integer part of } [n(1-\alpha)+1/2] \tag{3.5}$$

$$\eta_\alpha^{(3)} = (1-\delta)X_j + \delta X_{j+1}, \delta = 0 \; if \; g = 0; \; \delta = 1 \; if \; g > 0 \tag{3.6}$$

$$\eta_\alpha^{(4)} = \delta X_{j+1} + (1-\delta)(X_j + X_{j+1})/2, \quad \delta = 0 \; if \; g = 0; \; \delta = 1 \; if \; g > 0. \tag{3.7}$$

It is known that all of the above methods are asymptotically equivalent. Thus, if a large sample size is used (where for example M is determined from equation (3.1)), the choice of the *best* method is no longer critical. However, in an empirical study [37], it has been shown that $\eta_\alpha^{(4)}$ outperformed the other estimators when $g = 0$. It is noted that the methods based on the above estimators are restricted by the condition that $1 \leq n(1-\alpha) \leq n-1$. This implies that the smallest value of the false alarm probability α cannot be lower than $1/n$. Consequently, the threshold corresponding to the smallest false alarm probability which can be estimated by these procedures depends on the sample size. Thus, for a reasonable size of n, estimation of thresholds for very small false alarm probabilities cannot be made when these methods are used.

3.4.2 Estimates Motivated by the Extreme Value Theory

Extreme value distributions are obtained as limiting distributions of largest (or smallest) values of sample order statistics. Assuming independent trials, if $X_1 \leq X_2 \leq ... \leq X_n$ are order statistics from a common distribution function $F(x)$, then

the cumulative distribution function of the largest order statistic is given by

$$G_n(x) = P(X_n \leq x) = [F(x)]^n. \tag{3.8}$$

It is clear, as $n \rightarrow \infty$, that the limiting value of $G_n(x)$ approaches zero if $F(x)$ is less than 1 and unity if $F(x)$ is equal to 1 for a specified value of x. A standardized limiting distribution of X_n may be obtained by introducing the linear transformation, $a_n X_n + b_n$, where a_n and b_n are finite constants depending on the sample size n.

In Appendix A, using the theory of limiting distributions [45], it is shown that if there exist sequences a_n and b_n such that

$$\lim_{n\rightarrow\infty} P(\frac{X_n - b_n}{a_n} \leq x) = \lim_{n\rightarrow\infty} F^n(a_n x + b_n) = \lim_{n\rightarrow\infty} G_n(a_n x + b_n) \rightarrow \Lambda(x) \tag{3.9}$$

then the solution of equation (3.9) yields all the possible forms for the distribution function $G_n(x)$ in the limit as $n \rightarrow \infty$. The solutions to the above equation are derived in Appendix A and are rewritten here:

$$\Lambda(x) = exp(-e^{-x}) \quad x \geq 0 \tag{3.10}$$

$$\Lambda(x) = exp(-x^{-k}) \quad x \geq 0, \quad k > 0 \tag{3.11}$$

$$\Lambda(x) = exp(-(-x)^k) \quad x \leq 0, \quad k > 0. \tag{3.12}$$

In the limit, as n gets large, these are the three types of distribution functions to which the largest order statistic drawn from almost any smooth and continuous distribution function converge. By differentiating the three functions, we obtain analytical expressions for the limiting forms of the probability density functions. However, because of the differentiation, it should be recognized that these expressions may not be good approximations to the density functions. In practice, extreme value theory should always be applied to a distribution function, or equivalently, the area under the density function. For $x \geq 0$, differentiation of equations (3.10) and (3.11) result in

$$1. \quad \frac{d\Lambda(x)}{dx} \approx H(x) = e^{-x} \tag{3.13}$$

$$2. \quad \frac{d\Lambda(x)}{dx} \approx H(x) = kx^{-(k+1)} \quad k \geq 0. \tag{3.14}$$

The first equation above is the well known exponential distribution and the second equation is related to the Pareto distribution. The details that lead to the these analytical expressions are shown in Appendix A.

It remains to be explained how the distribution of the largest order statistic is related to the tails of the underlying PDF from which the samples are drawn. The relationship is based on the observation that inferences from short sequences are likely to be unreliable. In particular, instead of observing k sets of n samples and taking the largest order statistic from each of the k sets, it is better to observe a single set of nk samples and use the largest k samples from this set [46]. The k largest order statistics from a vector of nk observations constitute the tail of the underlying distribution especially when n is very large. Therefore, the limiting distribution of the largest order statistic closely approximates the tail of the underlying PDF for large n.

3.5 The Generalized Pareto Distribution

The Generalized Pareto Distribution (GPD) is defined for $x > 0$ by the distribution function

$$G(x) = 1 - (1 + \gamma x/\sigma)^{-1/\gamma}, \quad -\infty < \gamma < \infty, \sigma > 0, \gamma x > -\sigma. \tag{3.15}$$

This distribution has a simple closed form and includes a range of distributions depending upon the choice of γ and σ. For example, the exponential distribution results for $\gamma = 0$ and the uniform distribution is obtained when $\gamma = -1$. The GPD defined in equation (3.15) is valid for all $x > 0$ while equations (3.13) and (3.14) are valid only for large x.

The probability density function corresponding to the GPD is given by

$$g(x) = \frac{d}{dx}[1 - (1 + \frac{\gamma x}{\sigma})^{-1/\gamma}] = \frac{1}{\sigma}(1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}-1}. \tag{3.16}$$

If we let $\gamma \rightarrow 0$ in the above equation, note that

$$\lim_{\gamma \rightarrow 0} \frac{1}{\sigma}(1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}-1} = \frac{1}{\sigma} e^{-\frac{x}{\sigma}}. \tag{3.17}$$

Also, if we let x be large in equation (3.16), note that

$$\frac{1}{\sigma}(1+\frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}-1} \approx \frac{1}{\sigma}(\frac{\gamma}{\sigma})^{-\frac{1}{\gamma}-1}x^{-\frac{1}{\gamma}-1}. \tag{3.18}$$

Equations (3.17) and (3.18) are of the same form as equations (3.13) and (3.14). Thus, the GPD can be used to approximate both types of tail behavior exhibited by the right tail. Typical plots of the Generalized Pareto PDF for $\gamma < 0$ and $\gamma > 0$ are shown in Figures 3.2 & 3.3.

We wish to set thresholds for specified false alarm probabilities when the underlying density functions are unknown. To set very small false alarm probabilities, the tail of the PDF $f_{T_s}(t_s|H_0)$ has to be accurately modeled. Figure 3.4 represents a typical PDF of the test statistic with the tail region of the PDF being defined as that to the right of $t = t_0$. Figure 3.5 shows the tail translated to the origin. The choice for t_0 is somewhat arbitrary. For example, t_0 can be chosen such that the area in the shaded region equals 0.1, 0.05 or 0.01. It is the portion of the PDF to the right of t_0 that we are interested in modeling by the GPD. In particular, the tail region of the PDF is translated to the origin and modeled as a GPD. Once the estimates of σ and γ have been obtained, the GPD is multiplied by the area of the shaded region and translated back to the point t_0. In this way, the area under the PDF of the test statistic is maintained at unity.

3.5.1 Methods for Estimating the Parameters of the GPD

Suppose that the sample ordered statistics $X_1 \leq X_2 \leq ... \leq X_n$ are drawn from the distribution function $F(x)$. To estimate the right tail of this distribution it is necessary to determine a value (say x_0) and then use those sample observations which are greater than x_0 to obtain the quantity $z = x - x_0$. Once the tail observations have been chosen, the Generalized Pareto Distribution can be fitted to these observations by using standard methods of parameter estimation. Observe that the portion of the observations used from a complete set of samples depends on the choice of x_0. One approach to selecting x_0 is to make a histogram of the data set and choose x_0 to be near the point of inflection of the histogram. DuMouchel [47] proposed choosing x_0 to be the value such that $\int_{x_0}^{\infty} f_X(x)dx = 0.1$. Such an approach is less subjective and appears to be satisfactory for many applications. However, it is noted by DuMouchel that " using an even smaller fraction of observations would restrict profitable use of

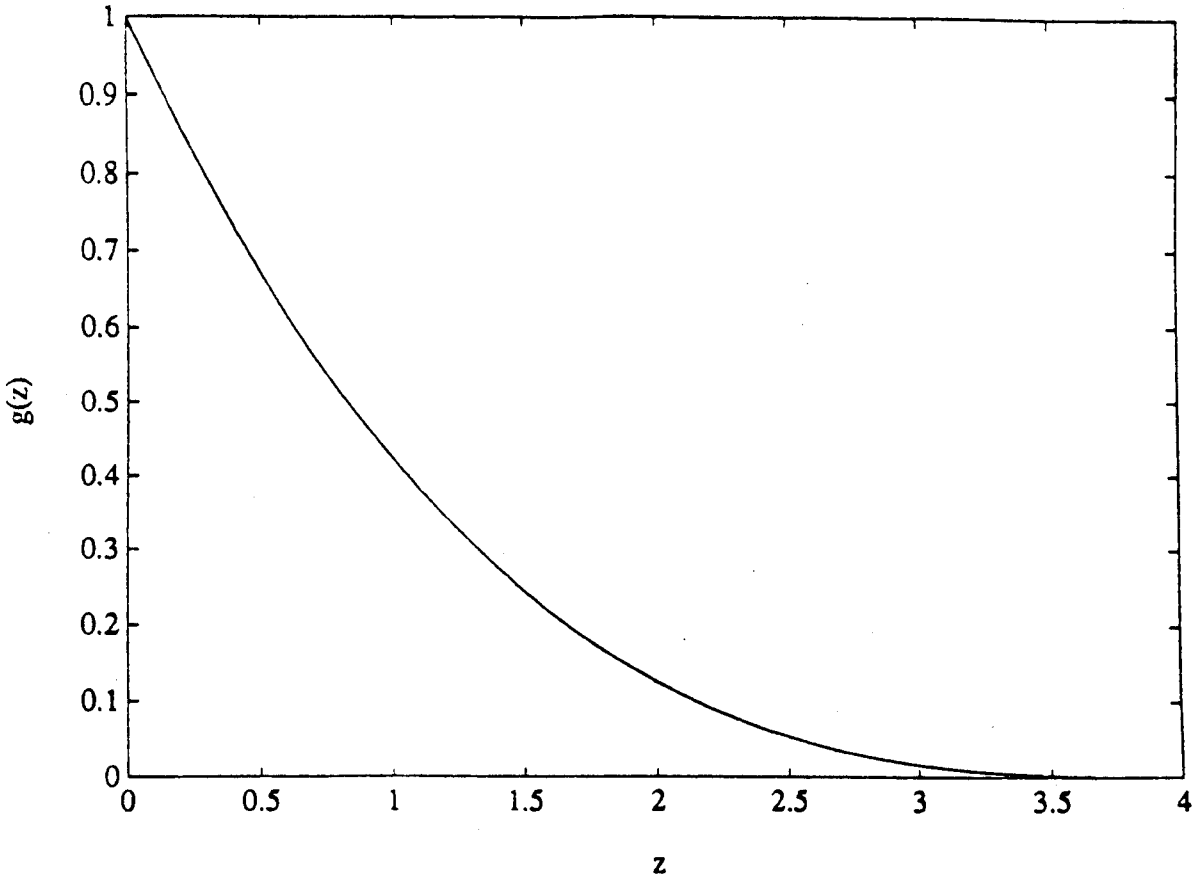


Figure 3.2: Generalized Pareto PDF, $\gamma = -1$.

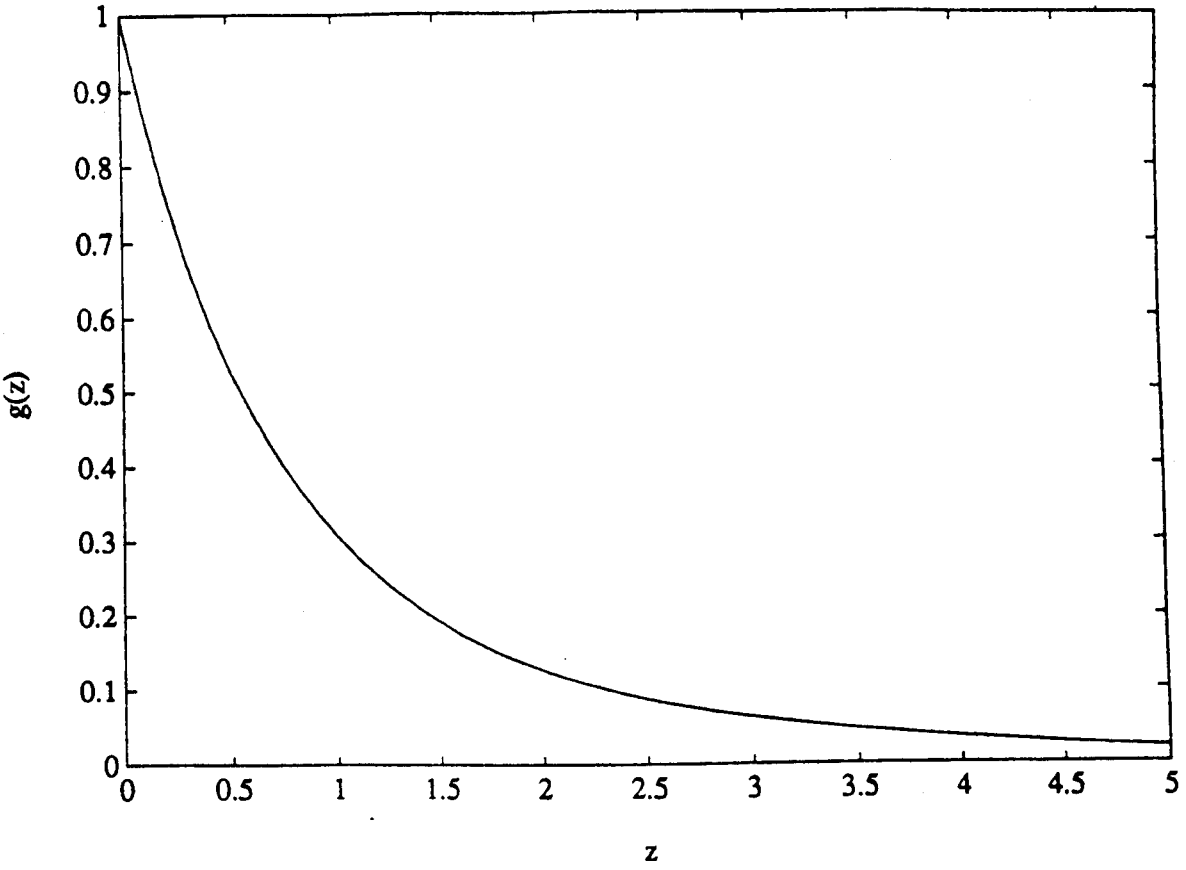


Figure 3.3: Generalized Pareto PDF, $\gamma = 1$.

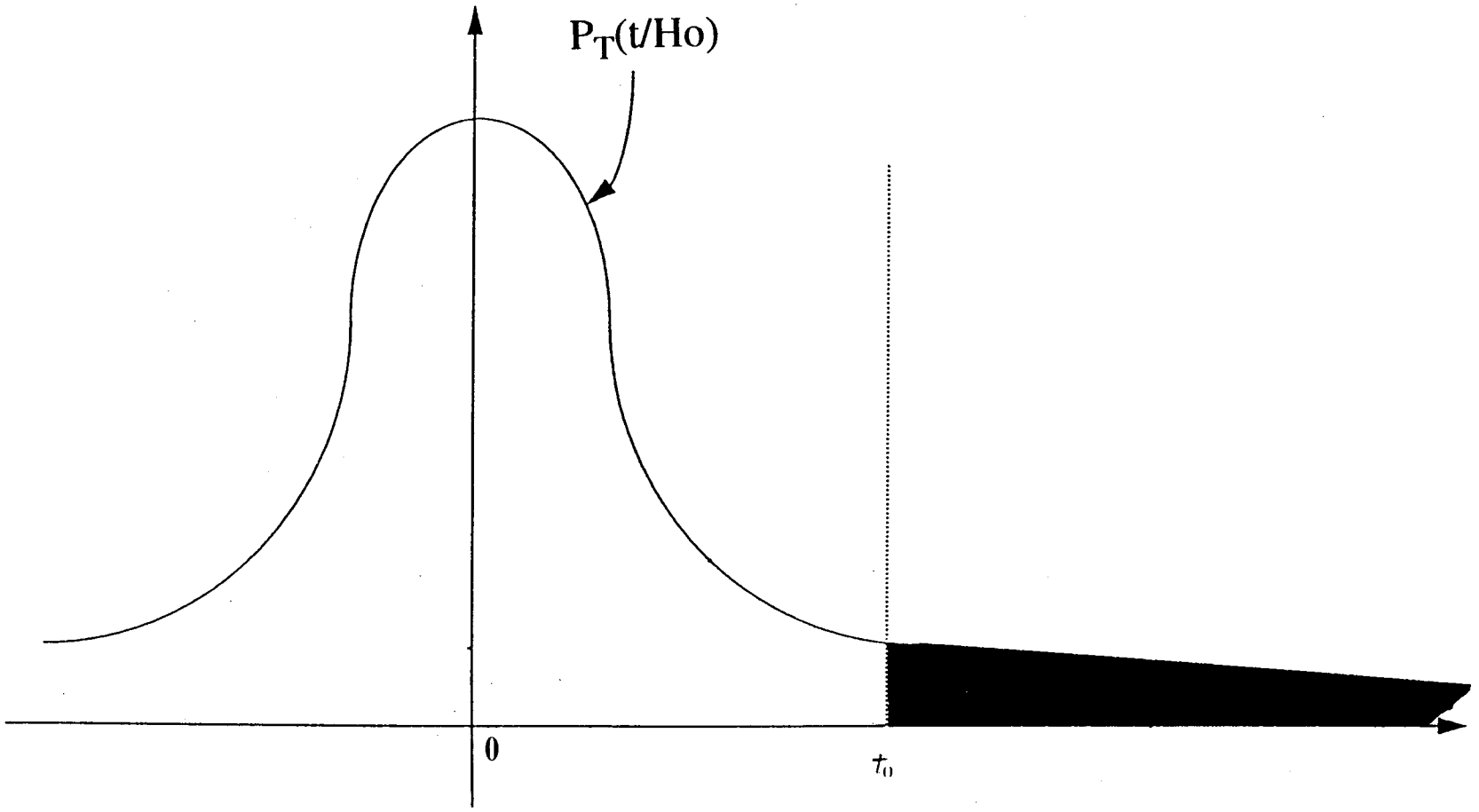


Figure 3.4: PDF of test statistic with tail region defined for $t \geq t_0$.

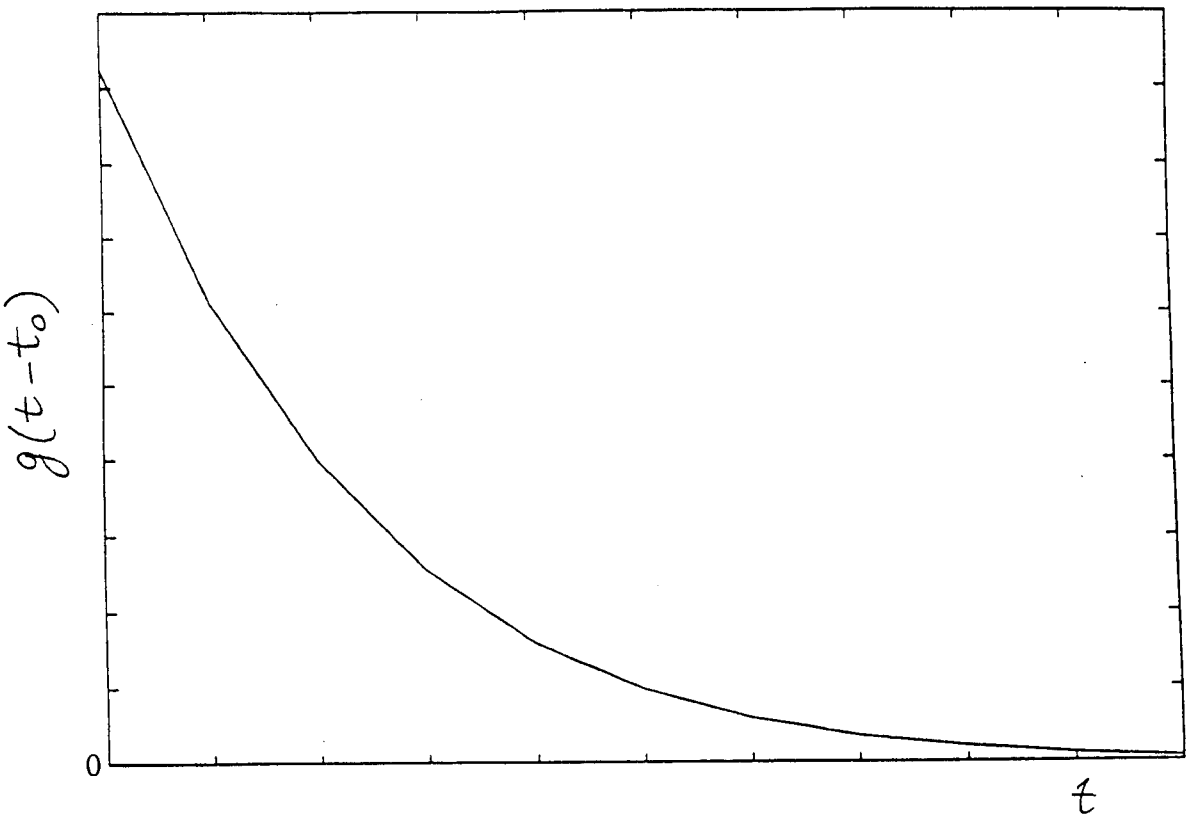


Figure 3.5: Tail of the test statistic shifted to origin.

the statistic to much larger sizes. On the other hand, to use more than the upper one tenth of a sample would seem to allow too much dependence on the central part of the distribution."In other words, if a smaller fraction is used, we need larger sample sizes to get an adequate number of samples for estimation and if a larger fraction is used, the body of the distribution may influence estimation of the tail.

Let x_0 be such that $1 - F(x_0) = \int_{x_0}^{\infty} f_X(x)dx = \lambda$. The distribution function to be used in approximating the tail can be written as

$$\tilde{F}(x) = (1-\lambda) + \lambda G(x - x_0) = 1 - \lambda[1 + \frac{\gamma}{\sigma}(x - x_0)]^{-1/\gamma} \quad x > x_0 \tag{3.19}$$

where $G(x)$ is given in equation (3.15). Assuming that the tail of a given distribution can be adequately approximated by equation (3.19), then the estimation problem of the distribution in the tail region is reduced to estimation of the parameters of the Generalized Pareto distribution.

In this chapter we consider three methods for the parameter estimation of the Generalized Pareto distribution. The three methods are maximum likelihood estimation, the method of probability weighted moments, and the ordered sample least squares approach. The first two methods, applied to the GPD, are discussed by Hosking and Wallis [31]. The ordered sample least squares approach is a new technique developed in this work. The performance of the three estimation procedures are compared on the basis of estimation bias and mean square error.

3.5.1.1 Maximum Likelihood Estimation

The probability density function corresponding to the GPD from equation (3.16), with x replaced by z, is

$$g(z) = \frac{1}{\sigma}(1 + \frac{\gamma z}{\sigma})^{-\frac{1}{\gamma}-1}. \tag{3.20}$$

Given a sample vector $[z_1, z_2, ..., z_m]$ from the GPD the joint density function $L_{\underline{Z}}(\underline{z})$ of the m samples, assuming independence, is given by

$$L_{\underline{Z}}(\underline{z}) = \frac{1}{\sigma^m}\prod_{i=1}^{m}(1 + \frac{\gamma z_i}{\sigma})^{-\frac{1}{\gamma}-1}. \tag{3.21}$$

To theoretically obtain the maximum likelihood estimates of σ and γ, the logarithm of the joint density function in equation (3.21) is differentiated with respect to σ and γ, respectively, and the derivatives are set to zero. Let the m largest observations

from the unknown distribution whose tail is being modeled by the GPD be placed in the vector $[x_{n-m+1}, x_{n-m+2}, ..., x_n]$. Translation of the tail region to the origin results in the vector $[x_{n-m+1} - x_0, x_{n-m+2} - x_0, ..., x_n - x_0] = [z_1, z_2, ..., z_m]$. Letting $\tau = \gamma/\sigma$ in equation (3.21) and differentiating the logarithm of the joint density function with respect to σ we get

$$\begin{aligned} -\frac{d}{d\sigma} log\ L_{\underline{Z}}(\underline{z}) &= \frac{d}{d\sigma}[m\ log\ (\sigma) + (1 + (\tau\sigma)^{-1}) \sum_{i=1}^{m} log(1 + \tau z_i)] \\ &= \frac{m}{\sigma} + (1 - \frac{1}{\sigma^2 \tau}) \sum_{i=1}^{m} log(1 + \gamma z_i/\sigma). \end{aligned} \tag{3.22}$$

By setting equation (3.22) to zero and solving for σ we obtain

$$\sigma(\tau) = \sum_{i=1}^{m} log(1 + \tau z_i)/(m\tau). \tag{3.23}$$

The expression for σ is now substituted into equation (3.22), so as to obtain a function of τ alone. $\hat{\tau}$ is derived by differentiating the expression

$$m\ log\ \sigma(\tau) + (1 + 1/(\sigma(\tau)\tau)) \sum_{i=1}^{m} log(1 + \tau z_i) \tag{3.24}$$

with respect to τ and setting the derivative equal to zero with the constraint that $\tau z_i > -1$. However, the differentiation leads to a nonlinear equation whose analytical solution is not known. This difficulty is circumvented by minimizing equation (3.24) numerically with respect to τ. The numerical minimization was performed using the Nelder-Mead algorithm [48]. Once the estimate for τ has been obtained, then $\hat{\sigma}$ is obtained from equation (3.23) and γ is estimated by $\hat{\gamma} = \hat{\sigma}\hat{\tau}$.

3.5.1.2 Probability Weighted Moments

The probability weighted moments of a continuous random variable Z with distribution function G are the quantities

$$M_{p,r,s} = E[Z^p G^r(Z)(1 - G(Z))^s] \tag{3.25}$$

where E is the expectation operator and p, r and s are real numbers. For the GPD it is convenient to choose $p = 1$ and $r = 0$, respectively. Then the probability weighted moments are

$$M_{1,0,s} = E[Z(1 - G(Z))^s]. \tag{3.26}$$

For the GPD there are two parameters to be estimated, σ and γ. Substituting $s = 0$ in equation (3.26), we get

$$\epsilon_0 = M_{1,0,0} = E[Z] = \int_0^\infty \frac{Z}{\sigma}(1 + \frac{\gamma Z}{\sigma})^{-\frac{1}{\gamma}-1} dZ. \tag{3.27}$$

Letting $1 + \frac{\gamma Z}{\sigma} = Y$, equation (3.27) results in

$$\begin{aligned} \epsilon_0 &= \frac{\sigma}{\gamma^2}\int_1^\infty (Y-1)Y^{-\frac{1}{\gamma}-1} dY \\ &= \frac{\sigma}{\gamma^2}[\frac{Y^{-\frac{1}{\gamma}-1}}{-\frac{1}{\gamma}+1} - \frac{Y^{-\frac{1}{\gamma}}}{-\frac{1}{\gamma}}]_1^\infty \\ &= \frac{\sigma}{1-\gamma}. \end{aligned} \tag{3.28}$$

Letting $s = 1$ in equation (3.26) we obtain

$$\epsilon_1 = M_{1,0,1} = E[Z(1-G(Z)] = \int_0^\infty \frac{Z}{\sigma}(1 + \frac{\gamma Z}{\sigma})^{-\frac{1}{\gamma}}(1 + \frac{\gamma Z}{\sigma})^{-\frac{1}{\gamma}-1} dZ. \tag{3.29}$$

Letting $1 + \frac{\gamma Z}{\sigma} = Y$, as before, equation (3.29) results in

$$\begin{aligned} \epsilon_1 &= \frac{\sigma}{\gamma^2}\int_1^\infty (Y-1)Y^{-\frac{2}{\gamma}-1} dY \\ &= \frac{\sigma}{\gamma^2}[\frac{Y^{-\frac{2}{\gamma}-1}}{-\frac{2}{\gamma}+1} - \frac{Y^{-\frac{2}{\gamma}}}{-\frac{2}{\gamma}}]_1^\infty \\ &= \frac{\sigma}{2(2-\gamma)}. \end{aligned} \tag{3.30}$$

The values of ϵ_0 and ϵ_1 are obtained from equations (3.28) and (3.30), respectively, for given values of σ and γ. Since there are two equations in two unknowns σ and γ can be obtained as functions of ϵ_0 and ϵ_1. Solving for σ and γ we obtain

$$\hat{\sigma} = 2\epsilon_0\epsilon_1/(\epsilon_0 - 2\epsilon_1) \tag{3.31}$$

and

$$\hat{\gamma} = 2 - \epsilon_0/(\epsilon_0 - 2\epsilon_1) \tag{3.32}$$

where ϵ_0 and ϵ_1 are estimated from the data by the estimators $\hat{\epsilon}_0 = \sum_{i=1}^m z_i/m$ and $\hat{\epsilon}_1 = \sum_{i=1}^m (m-i)z_i/\{m(m-1)\}$ [31]. Once the values of $\hat{\epsilon}_0$ and $\hat{\epsilon}_1$ are obtained the estimates of σ and γ are obtained by making use of equations (3.31) and (3.32). Note

that the method of probability weighted moments involves computationally simple expressions for the estimates.

3.5.1.3 The Ordered Sample Least Squares Method - A New Approach

The procedure used in maximum likelihood estimation is based on minimizing equation (3.24). The probability weighted moment estimates are obtained by equating the sample based averages with the theoretical values of the quantity $E[Z(1-G(Z))^s]$, s=0,1, where $Z = X - x_0$. A third approach is the ordered sample least squares method, which is based on the principle of minimizing the squared distance between the ordered sample and the expected value of the ordered sample. Computer simulations reveal that this can be a more suitable approach for estimating the parameters.

In Appendix A the method for evaluating the mean and the variance of the r^{th} order statistic from a sample of size n is presented. For the Generalized Pareto Distribution the mean and the variance of the r^{th} order statistic can be derived since the probability distribution function is known in closed form. Let x be replaced by z in equation (3.15) and let $G(z) = u$. Solution for z results in

$$z = G^{-1}(u) = \frac{\sigma}{\gamma}[(1-u)^{-\gamma} - 1]. \tag{3.33}$$

Making use of the above equation and equation (A.62) in Appendix A, the expected value of the r^{th} order statistic Z_r is

$$E(Z_r) = \frac{\sigma}{\gamma}\frac{n!}{(r-1)!(n-r)!}[\int_0^1((1-u)^{-\gamma}-1)u^{r-1}(1-u)^{n-r}du]. \tag{3.34}$$

The integral in the above equation can be broken into two parts as follows.

$$E(Z_r) = \frac{\sigma}{\gamma}\frac{n!}{(r-1)!(n-r)!}[\int_0^1(1-u)^{-\gamma}u^{r-1}(1-u)^{n-r}du - \int_0^1 u^{r-1}(1-u)^{n-r}du]. \tag{3.35}$$

From results presented in Gradshtyn and Ryzhik [49], the expression for $E(Z_r)$ becomes

$$\begin{aligned} E(Z_r) &= \frac{\sigma}{\gamma}\frac{n!}{(r-1)!(n-r)!}[\frac{(r-1)!(n-r-\gamma)!}{(n-\gamma)!} - \frac{(r-1)!(n-r)!}{n!}] \\ &= \frac{\sigma}{\gamma}[\frac{n!(n-r-\gamma)!}{(n-r)!(n-\gamma)!} - 1] \\ &= \frac{\sigma}{\gamma}[\frac{\Gamma(n+1)\Gamma(n-r-\gamma+1)}{\Gamma(n-r+1)\Gamma(n-\gamma+1)} - 1]. \end{aligned} \tag{3.36}$$

To calculate the variance of Z_r, we first calculate $E(Z_r^2)$. Making use of equation (3.33) and equation (A.65) in Appendix A, the expected value of Z_r^2 is

$$E(Z_r^2) = \frac{\sigma^2}{\gamma^2}\frac{n!}{(r-1)!(n-r)!}[\int_0^1((1-u)^{-\gamma}-1)^2u^{r-1}(1-u)^{n-r}du]. \quad (3.37)$$

Expanding the square in the integrand of the above equation gives

$$E(Z_r^2) = \frac{\sigma^2}{\gamma^2}\frac{n!}{(r-1)!(n-r)!}[\int_0^1((1-u)^{-2\gamma}-2(1-u)^{-\gamma}+1)u^{r-1}(1-u)^{n-r}du]. \quad (3.38)$$

Making use of results from [49], the above integral evaluates to

$$\begin{aligned} E(Z_r^2) &= \frac{\sigma^2}{\gamma^2}\frac{n!}{(n-r)!}[\frac{(n-r-2\gamma)!}{(n-2\gamma)!} - \frac{2(n-r-\gamma)!}{(n-\gamma)!} + 1] \\ &= \frac{\sigma^2}{\gamma^2}\frac{\Gamma(n+1)}{\Gamma(n-r+1)}[\frac{\Gamma(n-r-2\gamma+1)}{\Gamma(n-2\gamma+1)} - \frac{2\Gamma(n-r-\gamma+1)}{\Gamma(n-\gamma+1)} + 1]. \quad (3.39) \end{aligned}$$

From equations (3.36) and (3.39) and the relation $Var(Z_r) = E(Z_r^2) - E^2(Z_r)$, we have

$$\begin{aligned} Var(Z_r) &= \frac{\sigma^2}{\gamma^2}\frac{\Gamma(n+1)}{\Gamma(n-r+1)}[\frac{\Gamma(n-r-2\gamma+1)}{\Gamma(n-2\gamma+1)} \\ &- \frac{2\Gamma(n-r-\gamma)+1}{\Gamma(n-\gamma+1)} + 1] - \{\frac{\sigma}{\gamma}[\frac{\Gamma(n+1)\Gamma(n-r-\gamma+1)}{\Gamma(n-r+1)\Gamma(n-\gamma+1)} - 1]\}^2 \quad (3.40) \end{aligned}$$

Simplifying the above equation results in

$$Var(Z_r) = \frac{\sigma^2}{\gamma^2}[\frac{\Gamma(n+1)}{\Gamma(n-r+1)}\frac{\Gamma(n-r-2\gamma+1)}{\Gamma(n-2\gamma+1)} - \frac{\Gamma^2(n+1)}{\Gamma^2(n-r+1)}\frac{\Gamma^2(n-r-\gamma+1)}{\Gamma^2(n-\gamma+1)}]. \quad (3.41)$$

Letting $Q_r(\gamma) = \frac{\Gamma(n+1)}{\Gamma(n-r+1)}\frac{\Gamma(n-r-\gamma+1)}{\Gamma(n-\gamma+1)}$, results in

$$E(Z_r) = \mu_r = \frac{\sigma}{\gamma}\{Q_r(\gamma) - 1\} \quad (3.42)$$

$$Var(Z_r) = \sigma_r^2 = \frac{\sigma^2}{\gamma^2}\{Q_r(2\gamma) - (Q_r(\gamma))^2\}. \quad (3.43)$$

A computationally simpler expression can be found for $Q_r(\gamma)$ by making use of the

properties of the gamma function. Dividing $Q_r(\gamma)$ by $Q_{r-1}(\gamma)$ we get

$$\frac{Q_r(\gamma)}{Q_{r-1}(\gamma)} = \frac{\frac{\Gamma(n+1)}{\Gamma(n-r+1)}\frac{\Gamma(n-r-\gamma+1)}{\Gamma(n-\gamma+1)}}{\frac{\Gamma(n+1)}{\Gamma(n-r+2)}\frac{\Gamma(n-r-\gamma+2)}{\Gamma(n-\gamma+1)}} = \frac{n-r+1}{n-r-\gamma+1}. \tag{3.44}$$

Equation (3.44) reduces to

$$Q_r(\gamma) = \prod_{i=1}^{r} \frac{n-i+1}{n-i+1-\gamma}. \tag{3.45}$$

To find the least squares estimates of the parameters we write the following non-linear model for the r^{th} sample order statistic

$$Z_r = E(Z_r) + e_r, \quad r = 1, 2, ..., m \tag{3.46}$$

where the error term e_r has a distribution with mean 0 and variance σ_r^2. Since the order statistics are not independent, the errors are also not independent. Because of the nonlinear structure of the model in equation (3.46) and correlated errors, least squares estimation does not offer a straightforward solution to the estimation problem. Even so, in this study we proceed to use the ordered sample least squares (OSLS) procedure to estimate the parameters.

In equation (3.42), we note that the scale parameter σ appears linearly whereas the shape parameter γ does not. The least squares estimates are obtained by minimizing the quantity

$$S = \sum_{r=1}^{m} e_r^2 = \sum_{r=1}^{m} (Z_r - \sigma(Q_r(\gamma) - 1)/\gamma)^2. \tag{3.47}$$

Since σ appears linearly in the above expression, minimization can be achieved analytically. Differentiating equation (3.47) with respect to σ and setting the derivative equal to zero results in

$$2\sum_{r=1}^{m}((Z_r - \frac{\sigma}{\gamma}(Q_r(\gamma) - 1))((-\frac{1}{\gamma}(Q_r(\gamma) - 1)) = 0. \tag{3.48}$$

The solution for σ from the above equation is

$$\hat{\sigma}(\gamma) = \gamma \frac{\sum_{r=1}^{m} Z_r(Q_r(\gamma) - 1)}{\sum_{r=1}^{m}(Q_r(\gamma) - 1)^2}. \tag{3.49}$$

The expression for $\hat{\sigma}$ is substituted in equation (3.47) and the resulting expression

is minimized with respect to γ. After the substitution the resulting expression is nonlinear and minimization cannot be performed analytically. Using the Nelder-Mead algorithm [48], the minimization is done numerically. Once the estimate of $\hat{\gamma}$ is obtained, $\hat{\sigma}$ is obtained from equation (3.49).

Recall that the GPD is being used to approximate the tail of the underlying distribution. Hence, the ordered statistics Z_r, $r = 1, 2, ..., m$, from the GPD actually correspond to the ordered statistics $X_{n-m+1} - x_0$, $X_{n-m+2} - x_0 ... X_n - x_0$ from the underlying distribution.

The least squares procedure results in a computationally convenient algorithm. It is emphasized that the minimization of S is carried out only with respect to the single parameter γ. Furthermore, the underlying criterion is based on minimizing the distance between the empirical values and the expected values of the ordered samples. Some numerical comparisons are given in section 3.6.

3.5.2 Estimation of Thresholds

The Generalized Pareto Distribution that is estimated from the data is used to approximate the tail of the unknown, underlying distribution. We now show that the threshold is related to the approximating distribution function in a direct manner. With reference to equation (3.19), let $\hat{\eta}_\alpha$ denote the estimate of the threshold corresponding to a false alarm probability α. We then have

$$\tilde{F}(\hat{\eta}_\alpha) = 1 - \alpha = 1 - \lambda[1 + \frac{\gamma}{\sigma}(\hat{\eta}_\alpha - x_0)]^{-1/\gamma}. \tag{3.50}$$

Solution for $\hat{\eta}_\alpha$ results in

$$\hat{\eta}_\alpha = x_0 + \sigma(q^{-\gamma} - 1)/\gamma \tag{3.51}$$

where $\lambda = 1 - F(x_0)$, $q = (1 - \alpha)/\lambda$ and $x_0 = F^{-1}(1 - \lambda)$. For many applications DuMouchel [47] suggests that $\lambda = 0.1$ be used. As will be discussed in the subsequent sections, the optimal value of λ depends on the threshold being estimated. Since the distribution function $F(x)$ is not known, x_0 cannot be determined for a given value of α. Therefore, following common practice, the sample order statistic X_{n-m}, where $m = [\lambda n]$ and [.] denotes the integer part operator, is used as an estimate of x_0.

Distribution	OSLS	ML	PWM
Gaussian	-0.144	-0.151	-0.174
Weibull(3)	-0.163	-0.168	-0.194
Weibull(.67)	0.108	0.129	0.137
Weibull(.5)	0.201	0.265	0.263
Student-T(3)	0.290	0.260	0.261
Student-T(5)	0.132	0.099	0.090
Student-T(8)	0.031	0.006	-0.010
Lognormal(1)	0.232	0.259	0.258
Chi-square(1)	0.030	0.034	0.044
Chi-square(4)	-0.024	-0.033	-0.034
Chi-square(8)	-0.047	-0.058	-0.064

Table 3.1: Tail parameter γ describing the upper ten percent of various distributions.

3.6 Numerical Results

3.6.1 Characterization of Tail Shape for Known Distributions

We first discuss a method for estimating the parameters of the GPD when the underlying distribution is known. Choose x_0 such that $1 - F(x_0) = 0.1$. Then define the points p_i i=1,2,...1000 by

$$p_i = 0.90005 + 0.0001(i - 1). \tag{3.52}$$

Analytically evaluate $x_i = F^{-1}(p_i)$ from the known distribution. Using the 1000 values of x_i, the maximum likelihood estimation, the ordered sample least squares and the probability weighted moments procedures were applied to determine the corresponding γ values for various distributions. The results are given in Table 3.1. The number in parentheses for the Weibull and Lognormal distributions is the value of the shape parameter. For the remaining distributions the number denotes the degrees of freedom. Since σ is a scale parameter, the shape parameter γ best describes the tail shape. For the exponential and the uniform distributions the value of γ can be obtained analytically. $\gamma = 0$ for the exponential distribution and is -1 for the uniform distribution. Since the size of the tail decreases with decreasing γ, the relationship between the tail behavior and the corresponding values of the shape parameter γ can be clearly inferred from this table.

3.6.2 Empirical Properties of the Estimators for Known Distributions

Seven distributions with widely differing tail behaviors were chosen in order to investigate the adequacy of the approximation of extreme tails by the GPD and

to compare the three estimation procedures. The gamma distribution and Weibull distribution with shape parameter of value 3 have tails lighter than those of the exponential PDF. The tails of the chi-square distribution with 4 degrees of freedom and the student-T distribution with 8 degrees of freedom are approximately the same as those of the exponential PDF. Finally, the student-T distribution with 4 degrees of freedom and the Lognormal distribution with shape parameter of value 1 have tails heavier than those of the exponential PDF.

Let η and $\hat{\eta}$ denote the true and estimated thresholds, respectively. A Monte Carlo experiment was performed to investigate the normalized bias, $\frac{\hat{\eta}-\eta}{\eta}$ and the normalized mean square error $(\frac{\hat{\eta}-\eta}{\eta})^2$ of the proposed threshold estimates. The four sample sizes given by $m = 25, 50, 100$ and 1000 were considered. Each set of samples was obtained by generating n observations and taking the largest $m = 0.1n$ observations. For example, a set of samples of size 25 was obtained by selecting the largest 25 observations from a collection of 250 samples. For each of the four different values of m, k=200,000/m trials were performed for each of the seven distributions. The median of the normalized bias values was computed for each distribution and estimation procedure. The results for $P_F = 10^{-k}$, k=2,3,...7 are given in Table 3.2. Similarly the median of the positive square root of the normalized mean square error are presented in Table 3.3. The results in the two tables differ because the sign of $(\hat{\eta} - \eta)/\eta$ is lost in the normalized root mean square values computed in Table 3.3. Extremely poor estimates for η were obtained in a few of the trials. These poor estimates could severely influence an arithmetic mean of the estimates. To avoid this problem, median values were used in place of arithmetic means.

The empirical results in Table 3.2 indicate that the newly proposed ordered sample least squares estimator generally has a smaller normalized bias than the other estimators for small or moderate sample sizes. Overall the second smallest normalized bias is achieved by the probability weighted moments method. The maximum likelihood estimator has the largest normalized bias when $P_F \geq 10^{-5}$, especially for the long tailed distributions. The normalized bias of all three estimators decrease as the sample size increases. When the parent distribution is GPD,all three estimators perform very well. Even so, the ordered sample least square estimator outperforms the others. The relatively strong performance for the GPD is explained as follows. The extreme

value theory is based on the premise that tails of smooth continuous distributions tend towards the GPD. For the GPD, this premise is exactly satisfied. Hence, the corresponding performance is noticeably better than for other distributions.

The results for the median of the normalized root mean square error are surprising. The maximum likelihood estimator is known to be asymptotically efficient. This is always true when the samples are drawn from the underlying distribution (in our case from the generalized Pareto distribution). This property of the maximum likelihood estimator can be observed in Table 3.3 when m =1000 but not for smaller sample sizes. Although the ordered sample least squares method has a smaller normalized root mean square error in many cases, there is no clear winner with respect to this criterion.

From the empirical results which are based on a limited number of distributions and sample sizes, it is not easy to make a strong recommendation as to which method to use in practice. However, in terms of the normalized bias, the ordered samples least squares estimator appears to perform better than the other estimators in estimating the large thresholds when $P_F \leq 10^{-6}$. In any event, it is seen that the extreme value theory can be used successfully to determine threshold values, when the false alarm probability is very small.

Two practical advantages of estimation based on extreme value theory are: 1) When there is a constraint on the number of samples, the thresholds obtained from extreme value theory are expected to be closer to the true thresholds than those obtained by conventional Monte Carlo techniques. However, in both techniques an increase in sample size offers greater accuracy in estimating thresholds. 2) Because the estimate of the tail of the underlying distribution is in closed form, estimation can be made for thresholds corresponding to extremely small false alarm probabilities independent of the sample size. In experiments with fixed amounts of data, this is an important advantage.

3.6.3 Effect of the Choice of λ on the Threshold Estimates

As was mentioned previously, only those samples which exceed x_0 are used in estimating the GPD parameters. The value of x_0 is determined by λ. The results presented in Tables 3.2-3.3 were obtained by means of Monte Carlo experiments where $\lambda = 0.1$ was used independent of the value of P_F for which the threshold was being

estimated. When the false alarm probability was extremely small, the bias and root mean square errors were quite large for some distributions. The smaller the value of λ, the better will be the GPD approximation over the extreme tail being approximated. When λ is chosen too large, a better fit is found for that portion of the distribution closer to the center at the expense of lesser accuracy in the extreme tail. Of course, there is a tradeoff between the choice of λ and the number of data samples available for determining the parameters of the GPD.

In our application the major objective is to approximate the extreme tails corresponding to thresholds of 10^{-5} or smaller. Consequently, we explored the implications of selecting values less than 0.1 for λ. To accomplish this, we obtained the theoretical values of x_i for the standard Normal and Lognormal distributions corresponding to $F^{-1}(p_i)$ where $p_i = \frac{i-0.5}{n}$: i=1,2,...n, and $n = 1,000$ and $10,000$ respectively. These two distributions are chosen because they represent extremes: The Normal distribution is light tailed while the Lognormal is a heavy tailed distribution.

The number of the x_i samples used to determine the parameters of the GPD is given by λn. The parameters were estimated using the OSLS procedure for values of λ equal to 0.1, 0.05 and 0.01. The resulting GPDs were then used to determine the thresholds for false alarm probabilities given by $P_F = 10^{-k}$ where k=2,3,...7. These results are presented in Figure 3.6, where both the theoretical and approximated thresholds are plotted as a function of k for (A) Normal distribution (n=10,000), (B) Normal distribution (n=1000), (C) Lognormal distribution (n=10,000), (D) Lognormal distribution (n=1000). For $k \geq 5$, it is seen that $\lambda = 0.01$ (curve b) appears to be the best choice for approximating the thresholds. The best results were obtained with $n = 10,000$. However, good results were obtained with $n = 1,000$.

3.7 Examples

3.7.1 Known Distribution Case

To evaluate the accuracy of the threshold value estimates, 10000 random samples were generated from the Gaussian and Lognormal distributions and the upper tails of these two distributions were modeled as Generalized Pareto. In sections 3.4.1 and 3.4.3, theoretical values given by $x_i = F^{-1}(p_i)$ were used to estimate the tail. In this section randomly generated samples are used in place of the theoretical values. Choosing $\lambda = 0.01$, the theoretical thresholds of the Gaussian distribution for $P_F =$

$\lambda n = 25$

P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	-0.0112	0.0043	-0.0040	-0.0276	-0.0571	-0.087 2
Normal	ML	-0.0034	0.0187	0.0328	0.0358	0.0281	0.0137
Normal	PWM	-0.0084	-0.0208	-0.0560	-0.1015	-0.1464	-0.1924
Weibull(3)	OSLS	-0.0048	0.0013	-0.0041	-0.0202	-0.0418	-0.0619
Weibull(3)	ML	0.0039	0.0481	0.0938	0.1374	0.1776	0.2137
Weibull(3)	PWM	-0.0037	-0.0106	-0.0333	-0.0635	-0.0919	-0.1216
t(4)	OSLS	-0.0424	-0.0792	-0.1658	-0.2727	-0.3872	-0.4922
t(4)	ML	-0.0166	-0.1115	-0.2526	-0.4045	-0.5416	-0.6541
t(4)	PWM	-0.0218	-0.0929	-0.2160	-0.3498	-0.4761	-0.5881
t(8)	OSLS	-0.0221	-0.0186	-0.0572	-0.1164	-0.1975	-0.2879
t(8)	ML	-0.0104	-0.0468	-0.1169	-0.2077	-0.3055	-0.4033
t(8)	PWM	-0.0129	-0.0452	-0.1095	-0.2039	-0.3063	-0.4115
Chi-sq(4)	OSLS	-0.0209	-0.0039	0.0241	0.0333	-0.0088	-0.0104
Chi-sq(4)	ML	-0.0037	0.0943	0.2518	0.4571	0.6185	0.8810
Chi-sq(4)	PWM	-0.0144	-0.0205	-0.0334	-0.0576	-0.1254	-0.1624
Lognormal	OSLS	-0.0835	-0.0982	-0.0634	0.0016	0.1007	0.2567
Lognormal	ML	-0.0058	0.1836	0.5932	1.2736	2.4832	4.4947
Lognormal	PWM	-0.0543	-0.0878	-0.0931	-0.0728	-0.0228	0.0639
Pareto(-0.25)	OSLS	-0.0092	0.0208	0.0423	0.0631	0.0780	0.0874
Pareto(-0.25)	ML	-0.0030	0.0523	0.1190	0.1868	0.2479	0.2969
Pareto(-0.25)	PWM	-0.0077	0.0052	0.0121	0.0199	0.0237	0.0278

Table 3.2: Median of the normalized bias values for different percentiles. OSLS:Ordered Sample Least Square, ML:Maximum Likelihood, PWM:Probability Weighted Moments

		$\lambda n = 50$					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0036	0.0073	-0.0068	-0.0354	-0.0676	-0.1022
Normal	ML	0.0042	0.0323	0.0497	0.0578	0.0528	0.0380
Normal	PWM	-0.0012	-0.0118	-0.0459	-0.0861	-0.1318	-0.1742
Weibull(3)	OSLS	-0.0022	-0.0007	-0.0133	-0.0337	-0.0571	-0.0838
Weibull(3)	ML	0.0056	0.0500	0.0991	0.1436	0.1847	0.2199
Weibull(3)	PWM	-0.0014	-0.0105	-0.0342	-0.0629	-0.0937	-0.1256
t(4)	OSLS	-0.0147	-0.0646	-0.1800	-0.3209	-0.4501	-0.5063
t(4)	ML	-0.0068	-0.0867	-0.2264	-0.3736	-0.5120	-0.6291
t(4)	PWM	-0.0078	-0.0622	-0.1662	-0.2973	-0.4233	-0.5391
t(8)	OSLS	-0.0062	-0.0222	-0.0841	-0.1723	-0.2694	-0.3703
t(8)	ML	-0.0031	-0.0502	-0.1352	-0.2385	-0.3460	-0.4517
t(8)	PWM	-0.0032	-0.0336	-0.1064	-0.2041	-0.3051	-0.4046
Chi-sq(4)	OSLS	-0.0092	-0.0004	0.0051	0.0060	-0.0498	-0.0686
Chi-sq(4)	ML	0.0115	0.1134	0.2755	0.4775	0.6368	0.9150
Chi-sq(4)	PWM	-0.0041	-0.0087	-0.0191	-0.0407	-0.1123	-0.1488
Lognormal	OSLS	-0.0544	-0.0594	-0.0272	0.0458	0.1573	0.3274
Lognormal	ML	0.0092	0.2177	0.6336	1.3811	2.6197	4.7101
Lognormal	PWM	-0.0302	-0.0391	-0.0185	0.0413	0.1480	0.2977
Pareto(-0.25)	OSLS	-0.0052	0.0100	0.0214	0.0326	0.0404	0.0448
Pareto(-0.25)	ML	0.0005	0.0463	0.1011	0.1560	0.2003	0.2357
Pareto(-0.25)	PWM	-0.0050	-0.0018	-0.0012	-0.0019	-0.0023	-0.0012

Table 3.2: Median of the normalized bias values for different percentiles. (contd.)

		$\lambda n = 100$					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0017	-0.0016	-0.0253	-0.0637	-0.1040	-0.1464
Normal	ML	0.0068	0.0263	0.0306	0.0229	0.0063	-0.0185
Normal	PWM	0.0018	-0.0181	-0.0549	-0.1022	-0.1524	-0.1986
Weibull(3)	OSLS	0.0005	-0.0017	-0.0164	-0.0376	-0.0624	-0.0888
Weibull(3)	ML	0.0037	0.0270	0.0564	0.0840	0.1003	0.1158
Weibull(3)	PWM	0.0004	-0.0095	-0.0320	-0.0607	-0.0918	-0.1220
t(4)	OSLS	-0.0064	-0.0441	-0.1421	-0.2680	-0.3922	-0.5031
t(4)	ML	-0.0004	-0.0564	-0.1650	-0.2907	-0.4174	-0.5354
t(4)	PWM	-0.0003	-0.0478	-0.1403	-0.2636	-0.3809	-0.4949
t(8)	OSLS	-0.0024	-0.0134	-0.0751	-0.1606	-0.2578	-0.3548
t(8)	ML	0.0011	-0.0342	-0.1145	-0.2123	-0.3157	-0.4216
t(8)	PWM	0.0013	-0.0271	-0.0955	-0.1888	-0.2892	-0.3916
Chi-sq(4)	OSLS	-0.0032	-0.0028	-0.0077	-0.0198	-0.0841	-0.1111
Chi-sq(4)	ML	0.0175	0.1189	0.2655	0.4581	0.5917	0.8298
Chi-sq(4)	PWM	-0.0004	-0.0089	-0.0238	-0.0448	-0.1143	-0.1520
Lognormal	OSLS	-0.0159	-0.0542	-0.0876	-0.1089	-0.0940	-0.0617
Lognormal	ML	-0.0111	-0.0251	-0.0068	0.0536	0.1499	0.3104
Lognormal	PWM	-0.0165	-0.0210	0.0141	0.0924	0.2315	0.3965
Pareto(-0.25)	OSLS	-0.0023	0.0109	0.0255	0.0350	0.0419	0.0471
Pareto(-0.25)	ML	0.0033	0.0544	0.1170	0.1739	0.2215	0.2611
Pareto(-0.25)	PWM	-0.0014	0.0004	0.0052	0.0084	0.0112	0.0129

Table 3.2: Median of the normalized bias values for different percentiles. (contd.)

		$\lambda n = 1000$					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0035	-0.0013	-0.0244	-0.0613	-0.1010	-0.1432
Normal	ML	0.0059	0.0017	-0.0259	-0.0626	-0.1075	-0.1476
Normal	PWM	0.0028	-0.0192	-0.0586	-0.1064	-0.1560	-0.2016
Weibull(3)	OSLS	0.0013	-0.0023	-0.0175	-0.0381	-0.0627	-0.0885
Weibull(3)	ML	0.0020	-0.0018	-0.0159	-0.0386	-0.0641	-0.0909
Weibull(3)	PWM	0.0010	-0.0092	-0.0297	-0.0578	-0.0880	-0.1192
t(4)	OSLS	0.0058	-0.0044	-0.0605	-0.1574	-0.2690	-0.3715
t(4)	ML	0.0141	-0.0137	-0.1018	-0.2167	-0.3326	-0.4406
t(4)	PWM	0.0141	-0.0176	-0.1104	-0.2277	-0.3479	-0.4598
t(8)	OSLS	0.0033	-0.0021	-0.0452	-0.1167	-0.2001	-0.2896
t(8)	ML	0.0070	-0.0117	-0.0664	-0.1464	-0.2404	-0.3382
t(8)	PWM	0.0045	-0.0219	-0.0896	-0.1825	-0.2857	-0.3862
Chi-sq(4)	OSLS	0.0003	0.0012	-0.0057	-0.0167	-0.0826	-0.1107
Chi-sq(4)	ML	0.0012	-0.0021	-0.0152	-0.0354	-0.1026	-0.1349
Chi-sq(4)	PWM	0.0006	-0.0011	-0.0080	-0.0263	-0.0934	-0.1211
Lognormal	OSLS	-0.0038	-0.0221	-0.0259	0.0055	0.0646	0.1638
Lognormal	ML	-0.0098	0.0063	0.0616	0.1767	0.3495	0.5999
Lognormal	PWM	-0.0128	-0.0004	0.0567	0.1683	0.3400	0.5771
Pareto(-0.25)	OSLS	0.0002	0.0002	0.0012	0.0007	0.0003	0.0000
Pareto(-0.25)	ML	-0.0002	-0.0010	-0.0044	-0.0061	-0.0081	-0.0094
Pareto(-0.25)	PWM	0.0003	-0.0011	-0.0007	-0.0006	-0.0035	-0.0038

Table 3.2: Median of the normalized bias values for different percentiles.

		$\lambda n = 25$					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0558	0.1127	0.2022	0.2825	0.3507	0.4044
Normal	ML	0.0558	0.0909	0.1459	0.2057	0.2588	0.3070
Normal	PWM	0.0559	0.1215	0.2121	0.2920	0.3586	0.4117
Weibull(3)	OSLS	0.0257	0.0577	0.1089	0.1580	0.2031	0.2415
Weibull(3)	ML	0.0258	0.0531	0.0950	0.1378	0.1780	0.2139
Weibull(3)	PWM	0.0256	0.0624	0.1149	0.1659	0.2110	0.2495
t(4)	OSLS	0.1069	0.2261	0.4160	0.5989	0.7397	0.8405
t(4)	ML	0.1051	0.2353	0.4157	0.5812	0.7127	0.8097
t(4)	PWM	0.1019	0.2329	0.4213	0.5956	0.7368	0.8344
t(8)	OSLS	0.0781	0.1666	0.3073	0.4455	0.5701	0.6730
t(8)	ML	0.0779	0.1493	0.2554	0.3648	0.4689	0.5649
t(8)	PWM	0.0775	0.1752	0.3180	0.4544	0.5783	0.6787
Chi-sq(4)	OSLS	0.0610	0.1313	0.2441	0.3592	0.4650	0.5455
Chi-sq(4)	ML	0.0721	0.2179	0.4459	0.7901	1.1783	1.7789
Chi-sq(4)	PWM	0.0592	0.1384	0.2500	0.3622	0.4666	0.5446
Lognormal	OSLS	0.1335	0.2452	0.4362	0.6271	0.7785	0.8785
Lognormal	ML	0.1439	0.4007	0.7303	1.4149	2.7312	5.0774
Lognormal	PWM	0.1260	0.2582	0.4463	0.6281	0.7737	0.8705
Pareto(-0.25)	OSLS	0.0409	0.0787	0.1348	0.1752	0.2017	0.219
Pareto(-0.25)	ML	0.0402	0.0763	0.1419	0.2075	0.2640	0.3127
Pareto(-0.25)	PWM	0.0411	0.0866	0.1430	0.1817	0.2084	0.2240

Table 3.3: Median RMS errors for various percentiles. OSLS:Ordered Sample Least Square, ML:Maximum Likelihood, PWM:Probability Weighted Moments

$\lambda n = 50$

P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0401	0.0772	0.1391	0.1981	0.2548	0.3042
Normal	ML	0.0394	0.0689	0.1122	0.1559	0.1959	0.2328
Normal	PWM	0.0399	0.0865	0.1530	0.2192	0.2759	0.3273
Weibull(3)	OSLS	0.0180	0.0393	0.0743	0.1135	0.1511	0.1854
Weibull(3)	ML	0.0185	0.0509	0.0997	0.1447	0.1859	0.2214
Weibull(3)	PWM	0.0180	0.0442	0.0852	0.1263	0.1661	0.2017
t(4)	OSLS	0.0779	0.1826	0.3506	0.5179	0.6633	0.7724
t(4)	ML	0.0768	0.1910	0.3602	0.5244	0.6688	0.7762
t(4)	PWM	0.0760	0.1778	0.3332	0.4899	0.6303	0.7386
t(8)	OSLS	0.0561	0.1228	0.2316	0.3503	0.4666	0.5698
t(8)	ML	0.0553	0.1219	0.2226	0.3385	0.4504	0.5529
t(8)	PWM	0.0554	0.1306	0.2405	0.3613	0.4793	0.5807
Chi-sq(4)	OSLS	0.0431	0.0890	0.1678	0.2509	0.3351	0.4109
Chi-sq(4)	ML	0.0489	0.1661	0.3386	0.5487	0.7664	1.1112
Chi-sq(4)	PWM	0.0426	0.0939	0.1747	0.2584	0.3431	0.4185
Lognormal	OSLS	0.0975	0.1834	0.3439	0.5155	0.6660	0.7990
Lognormal	ML	0.0993	0.3381	0.6769	1.3921	2.6297	4.7240
Lognormal	PWM	0.0864	0.1954	0.3510	0.5143	0.6621	0.8012
Pareto(-0.25)	OSLS	0.0289	0.0534	0.0890	0.1162	0.1346	0.1486
Pareto(-0.25)	ML	0.0284	0.0602	0.1149	0.1675	0.2084	0.2417
Pareto(-0.25)	PWM	0.0293	0.0616	0.1032	0.1320	0.1533	0.1666

Table 3.3: Median RMS errors for various percentiles. (contd.)

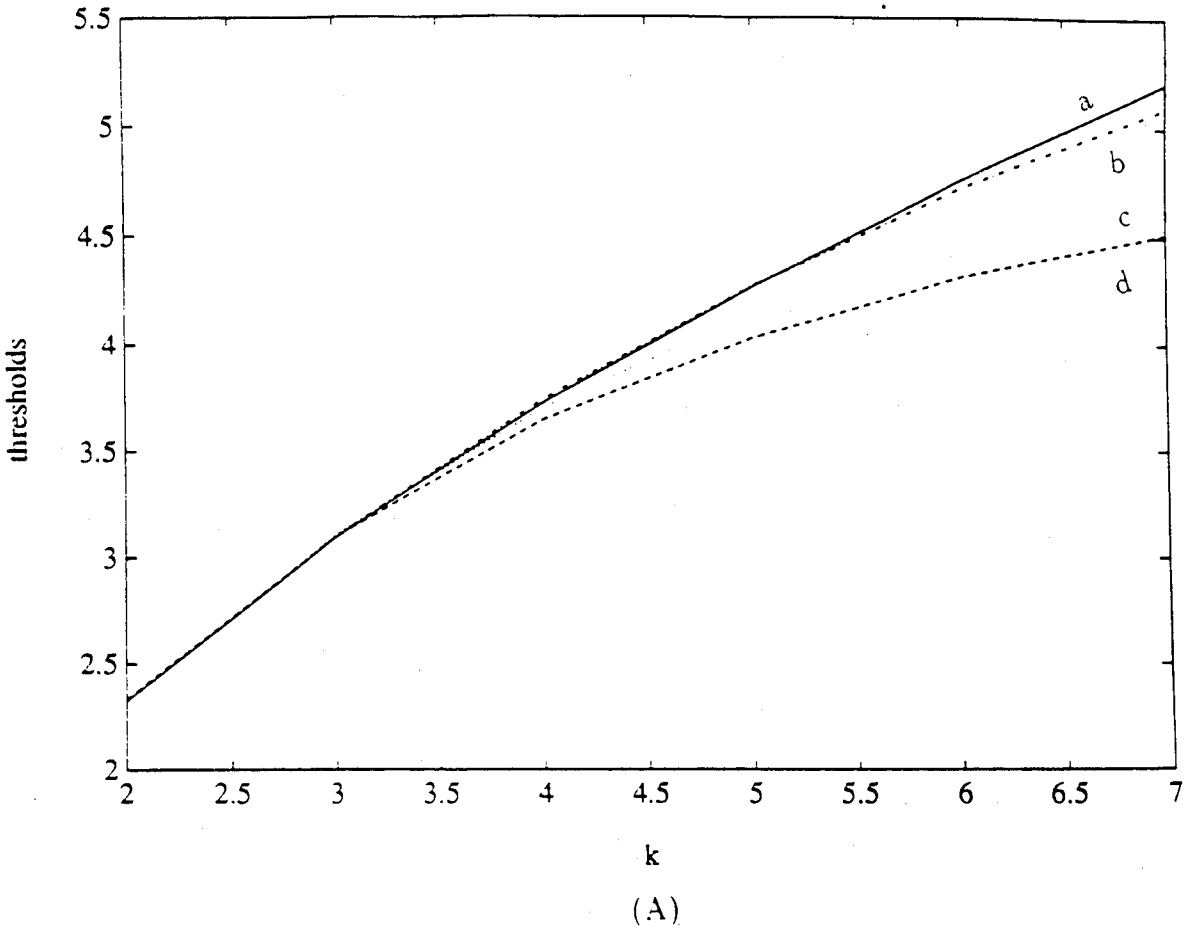
		$\lambda n = 100$					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0284	0.0522	0.0964	0.1414	0.1863	0.2305
Normal	ML	0.0290	0.0517	0.0840	0.1123	0.1433	0.1689
Normal	PWM	0.0281	0.0584	0.1090	0.1635	0.2134	0.2585
Weibull(3)	OSLS	0.0128	0.0273	0.0529	0.0811	0.1101	0.1378
Weibull(3)	ML	0.0131	0.0389	0.0790	0.1202	0.1570	0.1868
Weibull(3)	PWM	0.0126	0.0312	0.0622	0.0942	0.1284	0.1590
t(4)	OSLS	0.0550	0.1400	0.2801	0.4336	0.5739	0.6909
t(4)	ML	0.0525	0.1377	0.2716	0.4165	0.5497	0.6627
t(4)	PWM	0.0527	0.1334	0.2619	0.4046	0.5323	0.6469
t(8)	OSLS	0.0386	0.0914	0.1770	0.2761	0.3758	0.4732
t(8)	ML	0.0388	0.0896	0.1735	0.2710	0.3734	0.4763
t(8)	PWM	0.0384	0.0869	0.1727	0.2750	0.3777	0.4817
Chi-sq(4)	OSLS	0.0287	0.0649	0.1264	0.1932	0.2699	0.3373
Chi-sq(4)	ML	0.0350	0.1437	0.2959	0.4688	0.6092	0.8592
Chi-sq(4)	PWM	0.0283	0.0686	0.1289	0.1948	0.2730	0.3383
Lognormal	OSLS	0.0683	0.1527	0.2794	0.4174	0.5299	0.6290
Lognormal	ML	0.0652	0.1515	0.2690	0.4039	0.5465	0.6769
Lognormal	PWM	0.0647	0.1417	0.2519	0.3805	0.5218	0.6710
Pareto(-0.25)	OSLS	0.0201	0.0372	0.0637	0.0845	0.0997	0.1110
Pareto(-0.25)	ML	0.0197	0.0569	0.1192	0.1746	0.2221	0.2613
Pareto(-0.25)	PWM	0.0201	0.0434	0.0718	0.0952	0.1108	0.1220

Table 3.3: Median RMS errors for various percentiles. (contd.)

$\lambda n = 1000$

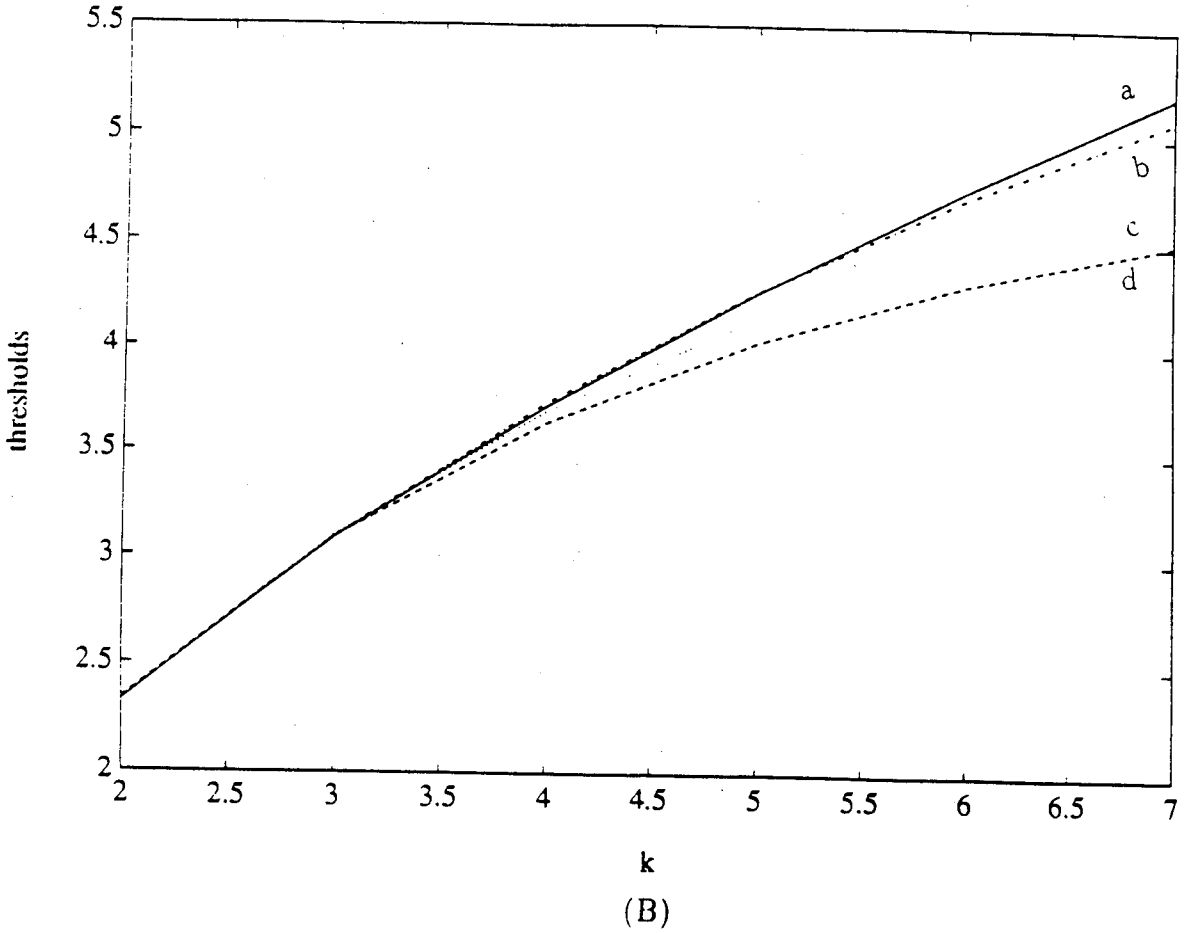
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0077	0.0182	0.0373	0.0643	0.1017	0.1440
Normal	ML	0.0087	0.0160	0.0362	0.0632	0.1075	0.1476
Normal	PWM	0.0081	0.0247	0.0586	0.1064	0.1560	0.2016
Weibull(3)	OSLS	0.0037	0.0086	0.0194	0.0393	0.0630	0.0890
Weibull(3)	ML	0.0040	0.0078	0.0191	0.0397	0.0649	0.0909
Weibull(3)	PWM	0.0036	0.0108	0.0300	0.0578	0.0880	0.1192
t(4)	OSLS	0.0203	0.0534	0.1383	0.2476	0.3717	0.4763
t(4)	ML	0.0213	0.0447	0.1083	0.2168	0.3326	0.4406
t(4)	PWM	0.0213	0.0499	0.1207	0.2306	0.3479	0.4598
t(8)	OSLS	0.0135	0.0298	0.0726	0.1379	0.2121	0.3018
t(8)	ML	0.0129	0.0272	0.0750	0.1518	0.2436	0.3406
t(8)	PWM	0.0129	0.0349	0.0939	0.1830	0.2863	0.3863
Chi-sq(4)	OSLS	0.0104	0.0207	0.0362	0.0588	0.1094	0.1408
Chi-sq(4)	ML	0.0099	0.0192	0.0363	0.0589	0.1095	0.1429
Chi-sq(4)	PWM	0.0100	0.0211	0.0400	0.0602	0.1103	0.1433
Lognormal	OSLS	0.0206	0.0528	0.1222	0.1836	0.2429	0.3276
Lognormal	ML	0.0195	0.0434	0.0984	0.2012	0.3581	0.5999
Lognormal	PWM	0.0201	0.0410	0.0927	0.1919	0.3445	0.5770
Pareto(-0.25)	OSLS	0.0061	0.0101	0.0158	0.0213	0.0247	0.0278
Pareto(-0.25)	ML	0.0063	0.0092	0.0154	0.0198	0.0243	0.0268
Pareto(-0.25)	PWM	0.0065	0.0126	0.0222	0.0306	0.0375	0.0428

Table 3.3: Median RMS errors for various percentiles.



(A)

Figure 3.6: Normal distribution, n=10,000 Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, ..., 7$. a:True, b:λ=0.01, c:λ=0.05, d:λ=0.10.



(B)

Figure 3.6: Normal distribution, n=1000 Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, ..., 7$. a:True, b:λ=0.01, c:λ=0.05, d:λ=0.10.

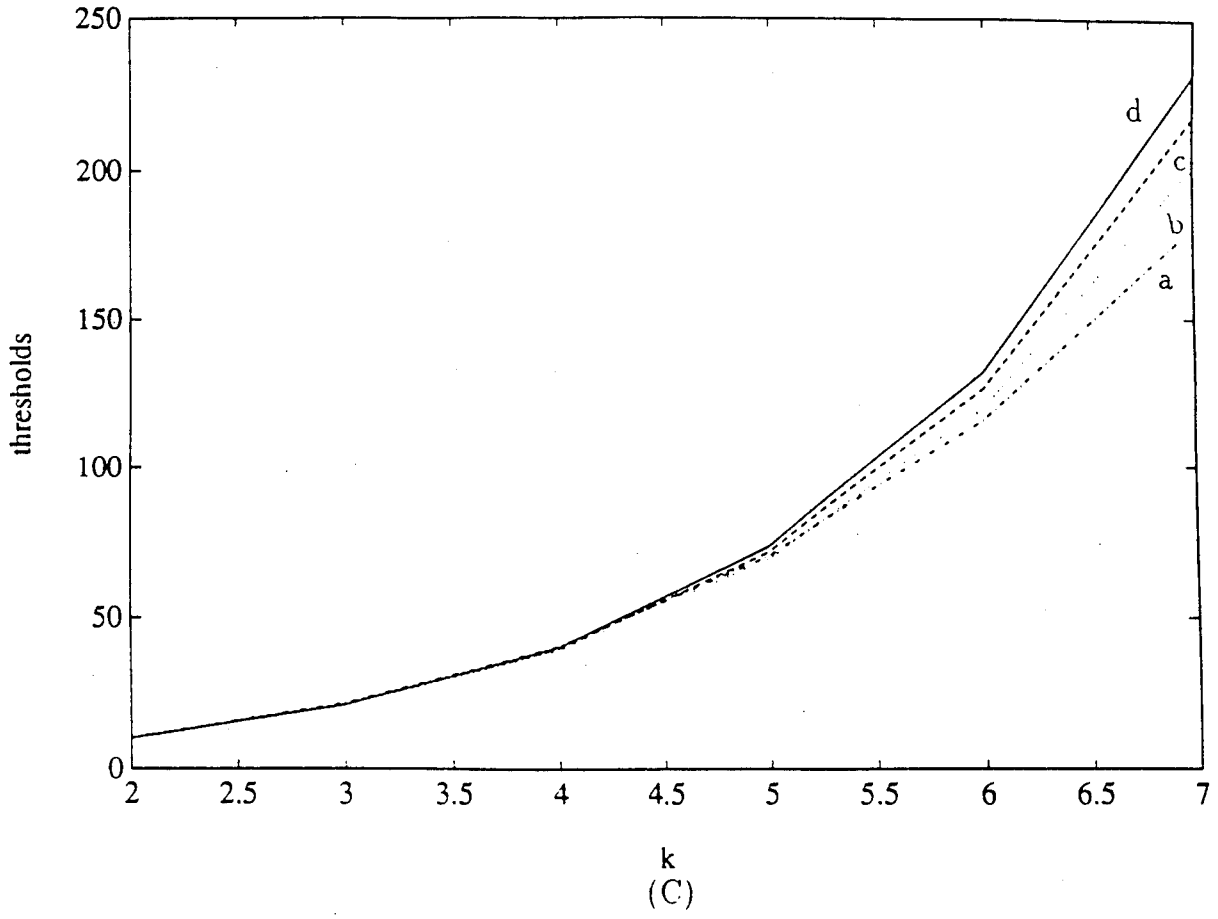


Figure 3.6: Lognormal distribution, n=10,000 Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, ..., 7$. a:True, b:λ=0.01, c:λ=0.05, d:λ=0.10.

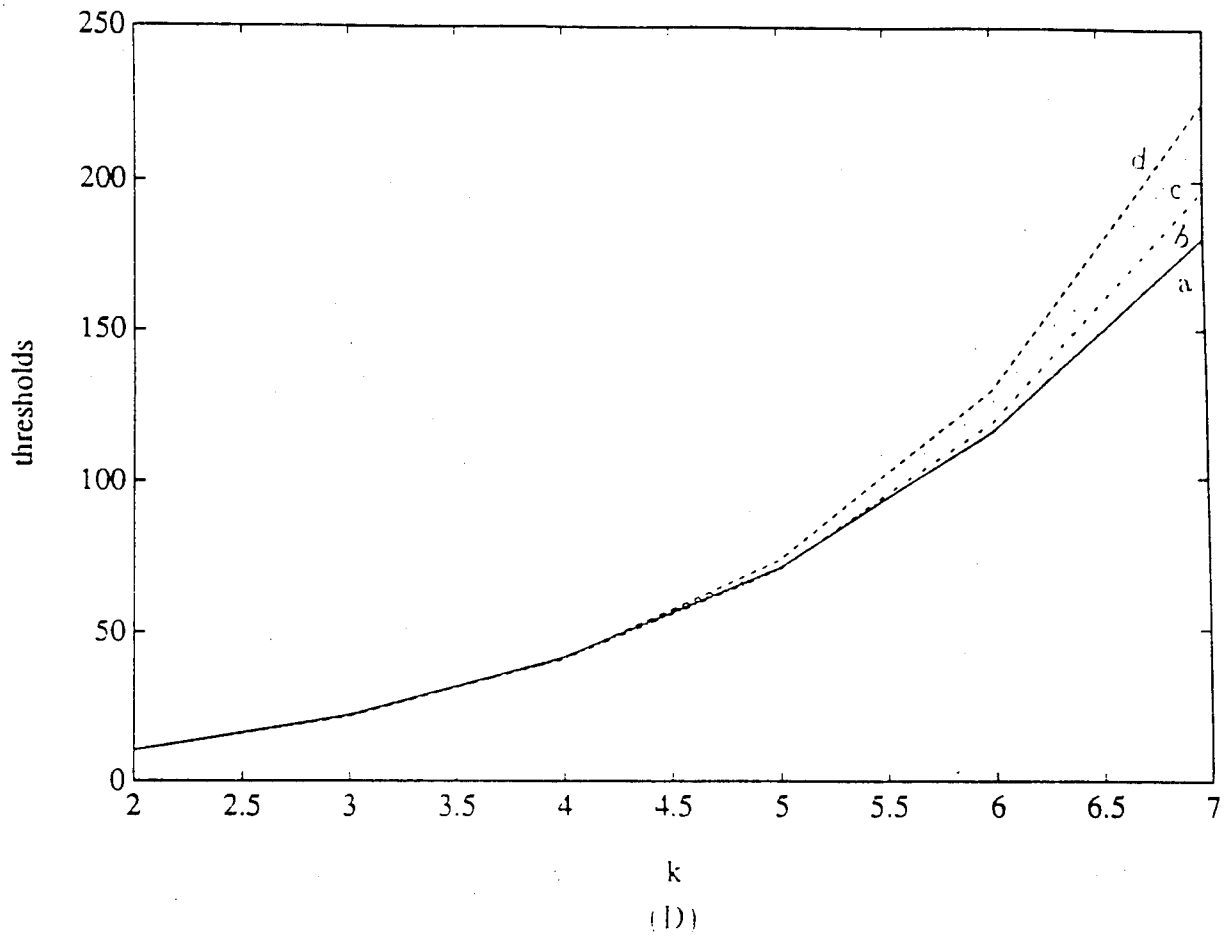


Figure 3.6: Lognormal distribution, n=1000 Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, ..., 7$. a:True, b:λ=0.01, c:λ=0.05, d:λ=0.10.

10^{-k} $k = 2, 3...7$ are 2.326, 3.090, 3.719, 4.265, 4.753 and 5.199, respectively. The thresholds estimated based on one set of random samples are 2.315, 3.223, 3.847, 4.370, 4.855 and 5.292. For the Lognormal distribution the theoretical thresholds corresponding to $P_F = 10^{-k}$ $k = 2, 3...7$ are 10.240, 21.982, 41.224, 71.157, 115.981 and 181.152. Once again, using λ=0.01, the thresholds estimated based on one set of random samples are 10.449, 22.862, 42.473, 69.216, 112.229 and 183.495. Note that the estimated results are very close to the true thresholds. We note here that these results were obtained on the basis of one set of observations from the two known distributions, corresponding to a particular seed value. For a different set of samples the estimates will be different depending on the tail behavior of that set of samples. But, unless the samples are really not a true representative of the distribution from which they are drawn, we expect that the estimates based on different samples should give threshold values that yield false alarm probabilities close to the design value.

3.7.2 An Unknown Distribution Case

In the previous section the underlying distributions were known to us and the estimates based on the extreme value theory were encouraging for both light and heavy tail behavior. In this example, we take a non-Gaussian problem where the underlying distribution is unknown.

The two hypotheses characterizing the detection problem are given in equations (2.1-2.2). We consider the weak signal case for which the clutter is much stronger than the background noise. The locally optimum detector (LOD) [8] has been shown to be suitable for the weak signal detection problem. Under hypothesis H_1, the signal is denoted by θs_i, where θ is a measure of the signal strength. For a deterministic signal and a given set of observations $\underline{r} = [r_1, r_2..., r_N]^T$ the LOD performs the ratio test

$$T_{LOD}(\underline{r}) = \frac{\frac{\partial P_{\underline{R}|H_1}(\underline{r}|H_1)}{\partial \theta}|\theta = 0}{\partial P_{\underline{R}|H_0}(\underline{r}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_k \tag{3.53}$$

where $P_{\underline{R}|H_i}(\underline{r}|H_i)$ is the joint PDF of $r_1, r_2, ...r_N$ under hypothesis H_i: i=0,1.

Martinez, Swaszek and Thomas [22] studied the locally optimal detection problem for non-Gaussian distributions and considered the bivariate Laplace distribution as an example. In this section we illustrate the procedure for determining the thresholds of a LOD based on N=2 and the received samples under H_0 having the bivariate

Laplace distribution given by

$$f_{\underline{R}}(r_1, r_2) = \frac{1}{2\pi |M|^{1/2}} K_0[(2\underline{r}^T M^{-1} \underline{r})^{1/2}] \tag{3.54}$$

where M is the covariance matrix for the two samples, $|M|$ denotes its determinant, $\underline{r}^T M^{-1} \underline{r}$ is equal to $(r_1^2 - 2\rho r_1 r_2 + r_2^2)/(1-\rho^2)$, ρ is the correlation coefficient between R_1 and R_2 and $K_0(.)$ is the modified Bessel function of the second kind of zero order. The resulting locally optimum detector statistic is [22]

$$T_{LOD}(r_1, r_2) = (\frac{2}{\underline{r}^T M^{-1} \underline{r}})^{1/2} \frac{K_1[(2\underline{r}^T M^{-1} \underline{r})^{1/2}]}{K_0[(2\underline{r}^T M^{-1} \underline{r})^{1/2}]} \times \underline{s}^T M^{-1} \underline{r} \tag{3.55}$$

where $\underline{s} = (s_1, s_2)^T$, $\underline{s}^T M^{-1} \underline{r} = (r_1 - \rho r_2)s_1 + (r_2 - \rho r_1)s_2$ and $K_1(.)$ is the modified Bessel function of the second kind of first order. s_1 *and* s_2 are the known signal levels. In this example we take $s_1 = 1$ and $s_2 = -1$. Because of the complexity of $T_{LOD}(.)$, it is not possible to determine a closed form expression for its probability density function.

In many applications in radar, thresholds have to be set to achieve desired false alarm probabilities based on a sample size which is orders of magnitude less than $10/P_F$. As will be pointed out later, the statistic in equation (3.55) represents a worst case situation in the sense that our simulations indicate that the variance of the test statistic is extremely large. To investigate the reliability of the thresholds estimated based on extreme value theory with smaller sample sizes, 10,000 pairs of observations (r_1, r_2) were generated from the bivariate Laplace distribution given in equation (3.54), with $\rho = 0.90$. The values of $T_{LOD}(r_1, r_2)$ were computed for each pair and sorted in increasing order. Corresponding to $\lambda = 0.01$, the largest 100 values of the underlying statistic (the top one per cent) were selected to fit the Generalized Pareto Distribution. This experiment was repeated 250 times. The threshold corresponding to a certain false alarm probability P_F of the distribution of the statistic $T_{LOD}(r_1, r_2)$ is estimated from equation (3.51) as $\tilde{\eta}_\alpha = x_0 + \hat{\sigma}[(\frac{1-\alpha}{0.01})^{-\hat{\gamma}} - 1]/\hat{\gamma}$ where x_0 is the 9900^{th} largest value of the statistic. Thresholds were estimated for false alarm probabilities $P_F = 10^{-k}$, $k = 2, ..., 7$ for each repetition of the experiment. Histograms of these threshold values are shown in figure 3.7, for the different P_Fs. To give a better appreciation for the range of values, the bins are not necessarily of

equal width. The histograms give an indication of the spread in the threshold values depending on the particular samples collected. From the histograms corresponding to false alarm probabilities of 10^{-2}, 10^{-3} and 10^{-4} we can see that the threshold estimates obtained on the basis of even one set of samples is likely to approximately yield the desired P_F. Since the underlying distribution of $T_{LOD}(.)$ is unknown, one measure of the accuracy of the estimate is the extent to which most of the estimates fall in one bin of the histogram. Also, we can see that there is negligible overlap between the estimated threshold values in the histograms for the three different values of P_F. This supports the claim that the estimated threshold is likely to yield a false alarm probability which is of the same order as the desired P_F. There is a higher overlap in the thresholds of the histograms for P_F=10^{-5}, 10^{-6} and 10^{-7}. Also, there is much higher spread in the threshold values estimated. Based on the excellent results obtained for the same choices of P_F in the known cases of the previous section, these results are surprising. However, it is explained as follows. The γ values of the GPD estimated for the different repetitions of this experiment lie in the range $0.45 - 0.55$. This represents an extremely heavy tailed distribution. From Table 3.1 we see that the Lognormal distribution, which is quite a heavy tailed distribution, has γ=0.232. The heavy tailed nature of the detector statistic can also be observed by comparing the large threshold values seen in the histograms with the corresponding thresholds of the Gaussian and the Lognormal distributions. The variance of the GPD is given by

$$\begin{aligned} Var(X) &= \frac{\sigma^2}{(1-\gamma^2)(1-2\gamma)} && \gamma < 0.5 \\ &= \infty && \gamma \geq 0.5 \end{aligned} \tag{3.56}$$

Thus, the bivariate Laplace results in a very highly fluctuating statistic with an extremely large variance. As such, it represents a 'worst case' situation for empirically determining the threshold. A much larger sample size is needed to obtain reliable threshold estimates because of the exceedingly large tail of the underlying distribution.

In general, an indication of how heavy the true tail may be for an unknown distribution is given by the estimate of γ for the GPD. When an extremely heavy tail is indicated, another strategy for estimating the thresholds when P_F is very small is to

choose the median value of the thresholds estimated when the experiment is repeated a specified number of times with 10,000 samples in each repetition. The choice of the median as the estimator ensures that very large and very small values do not affect the results. For the present example, we chose to repeat the 250 trials three times. By counting the number of estimates that fell into the bins centered at 20, 28 and 36 for P_F=10^{-5}, 40, 50,70 and 90 for P_F=10^{-6} and 100 and 150 for P_F=10^{-7}, it was found that 88 percent of the estimates fell into these bins. Thus, even for this extremely large tailed example, we believe that use of the GPD has allowed us to estimate useful values for the thresholds with sample sizes much smaller than $10/P_F$.

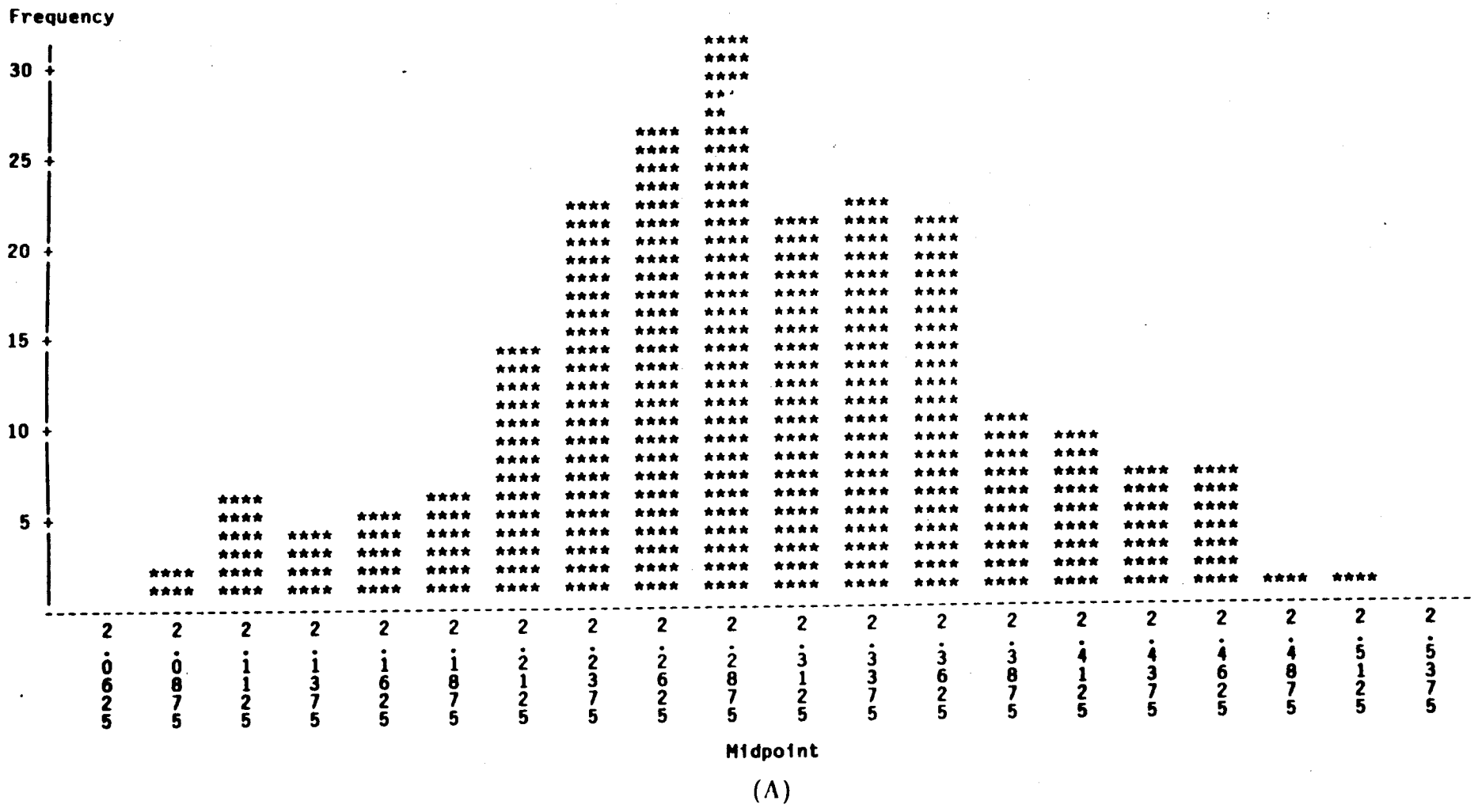
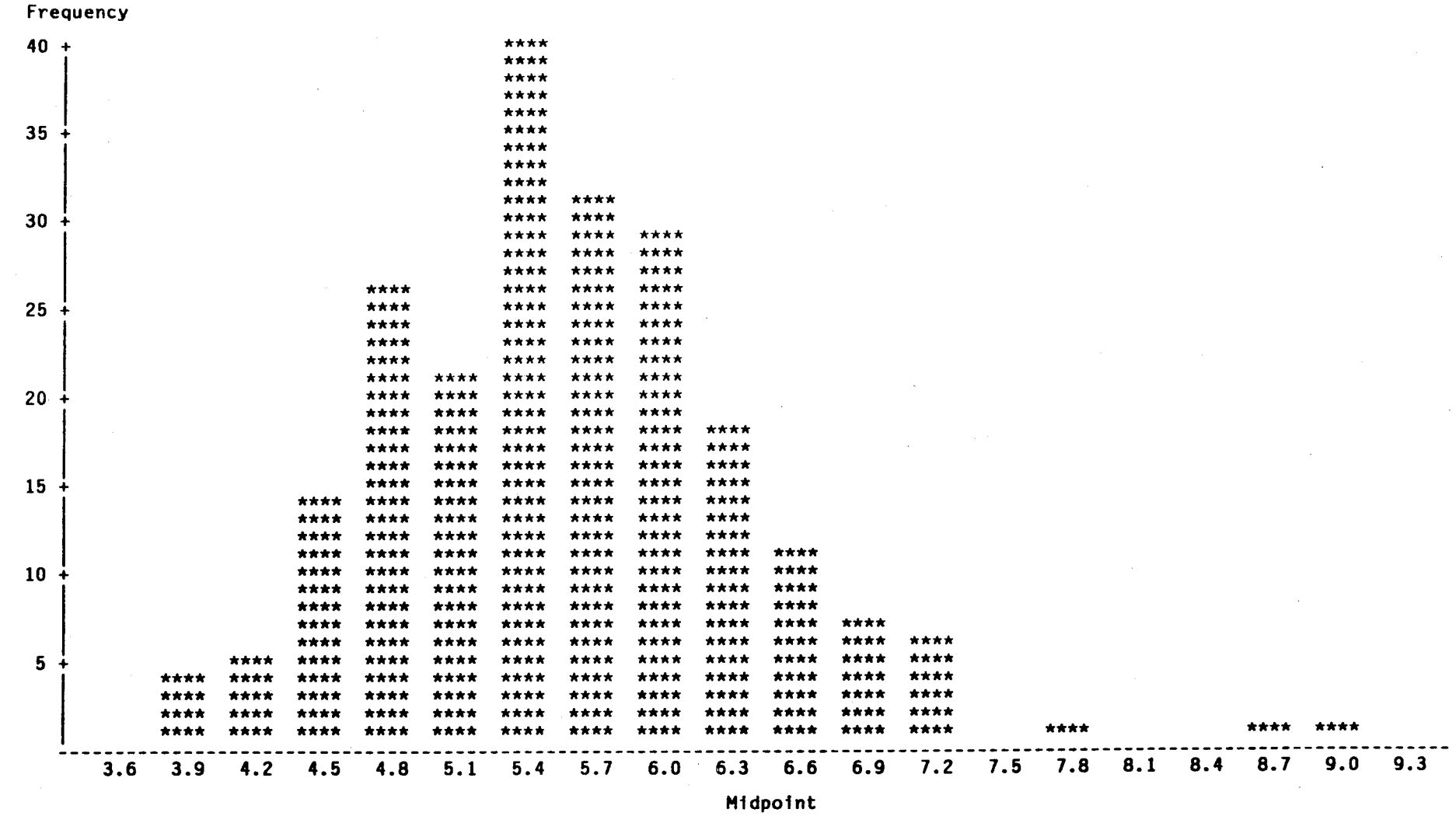
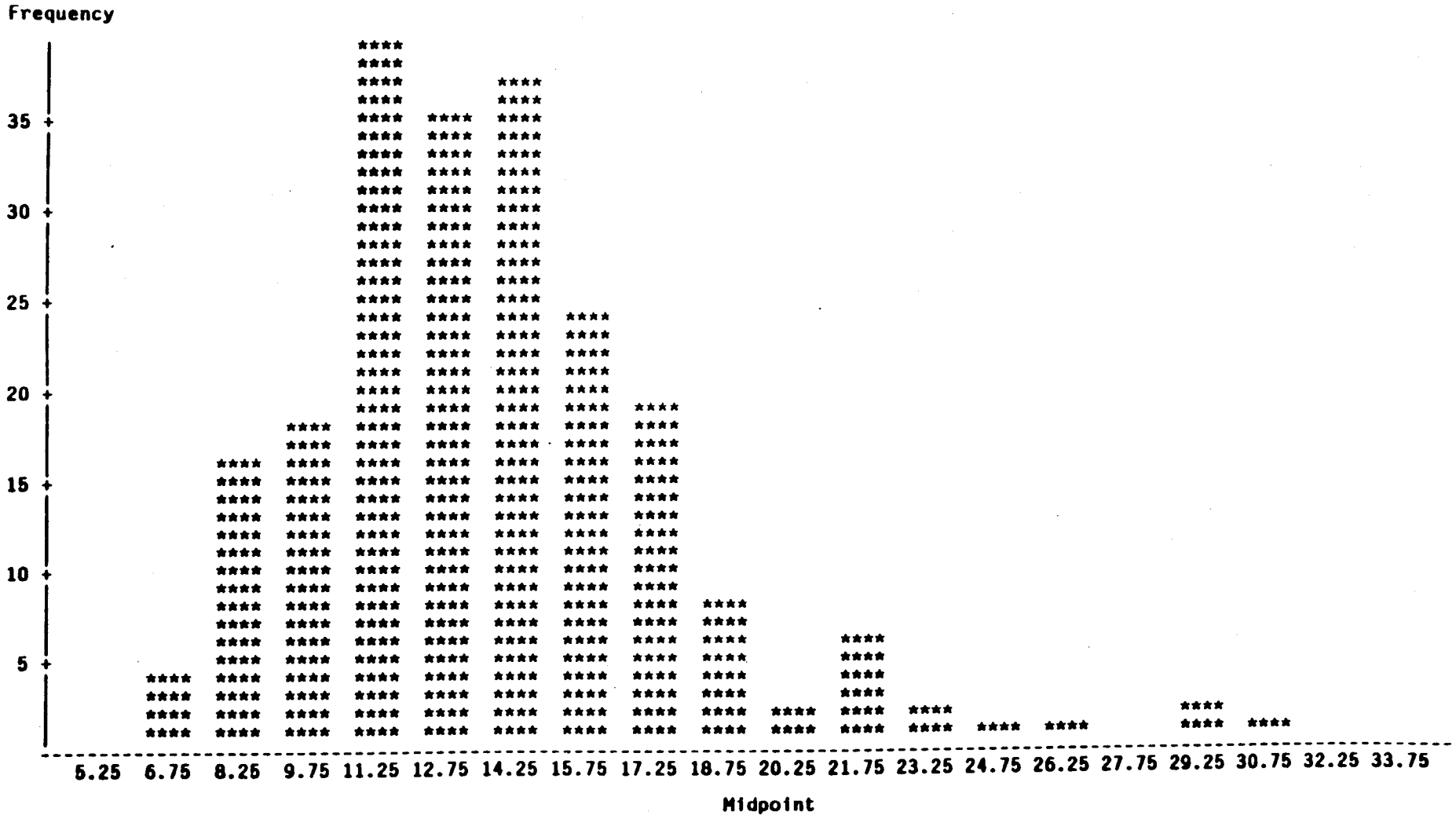


Figure 3.7: Histograms of threshold values. (A) $P_F = 10^{-2}$ (B) $P_F = 10^{-3}$ (C) $P_F = 10^{-4}$ (D) $P_F = 10^{-5}$ (E) $P_F = 10^{-6}$ (F) $P_F = 10^{-7}$



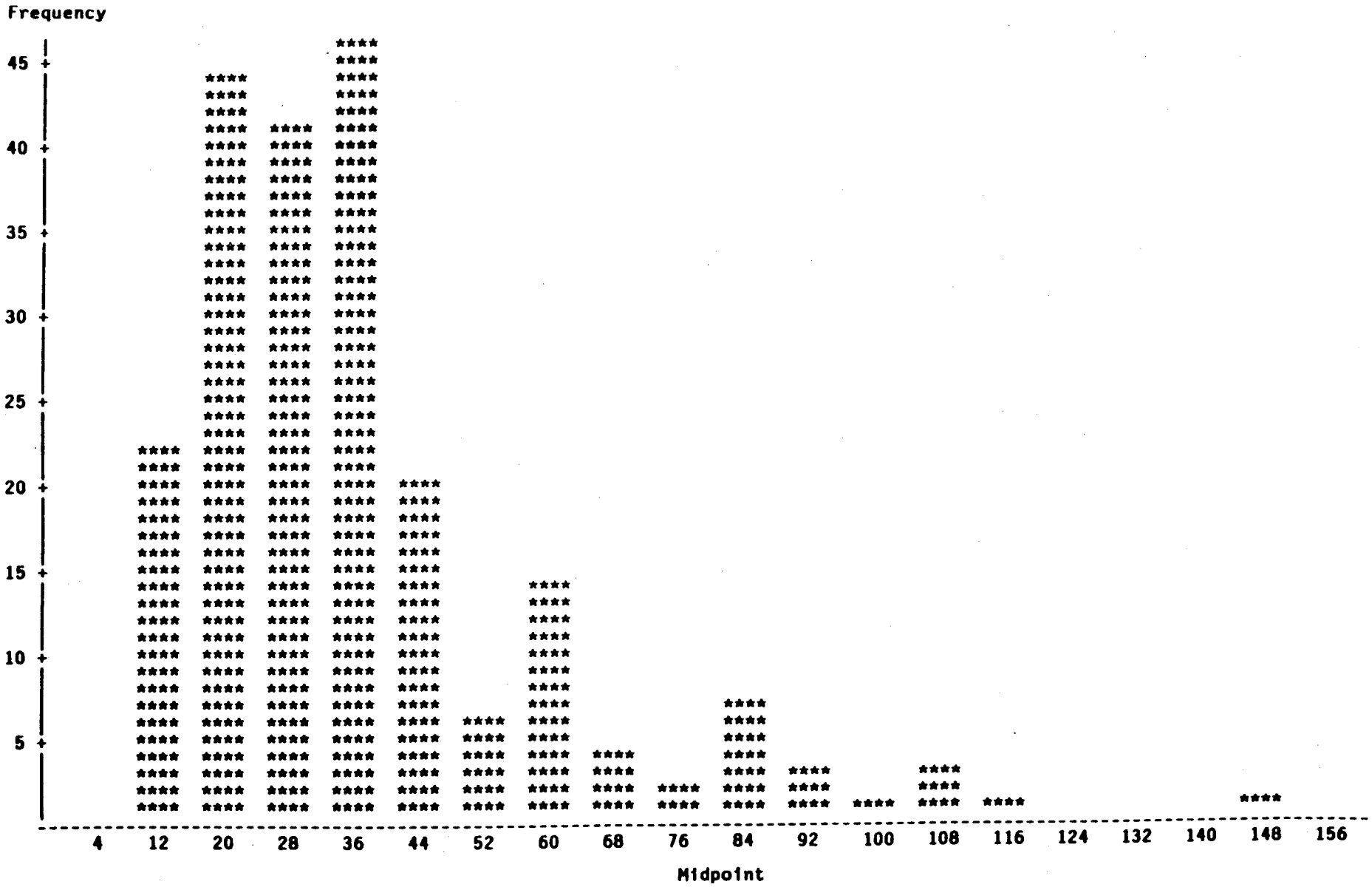
(B)

Figure 3.7: Fig 3.7 Contd.

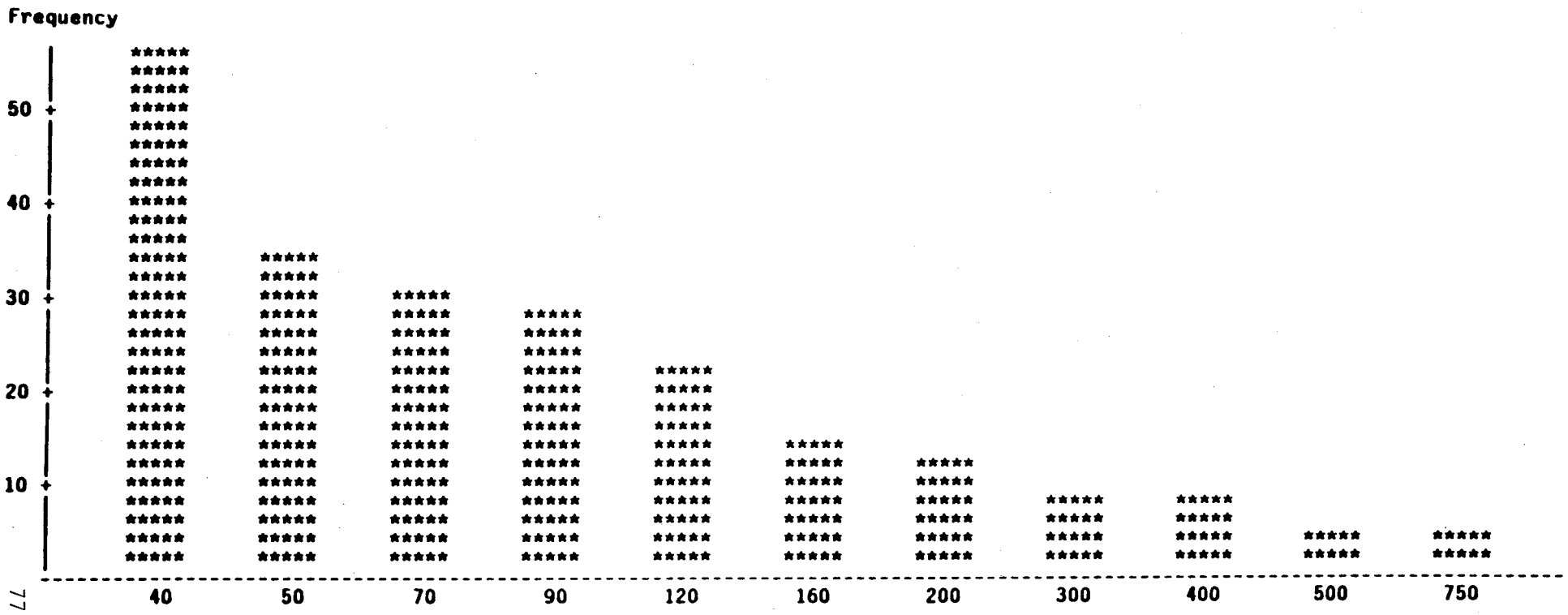


(C)

Figure 3.7: Fig 3.7 Contd.

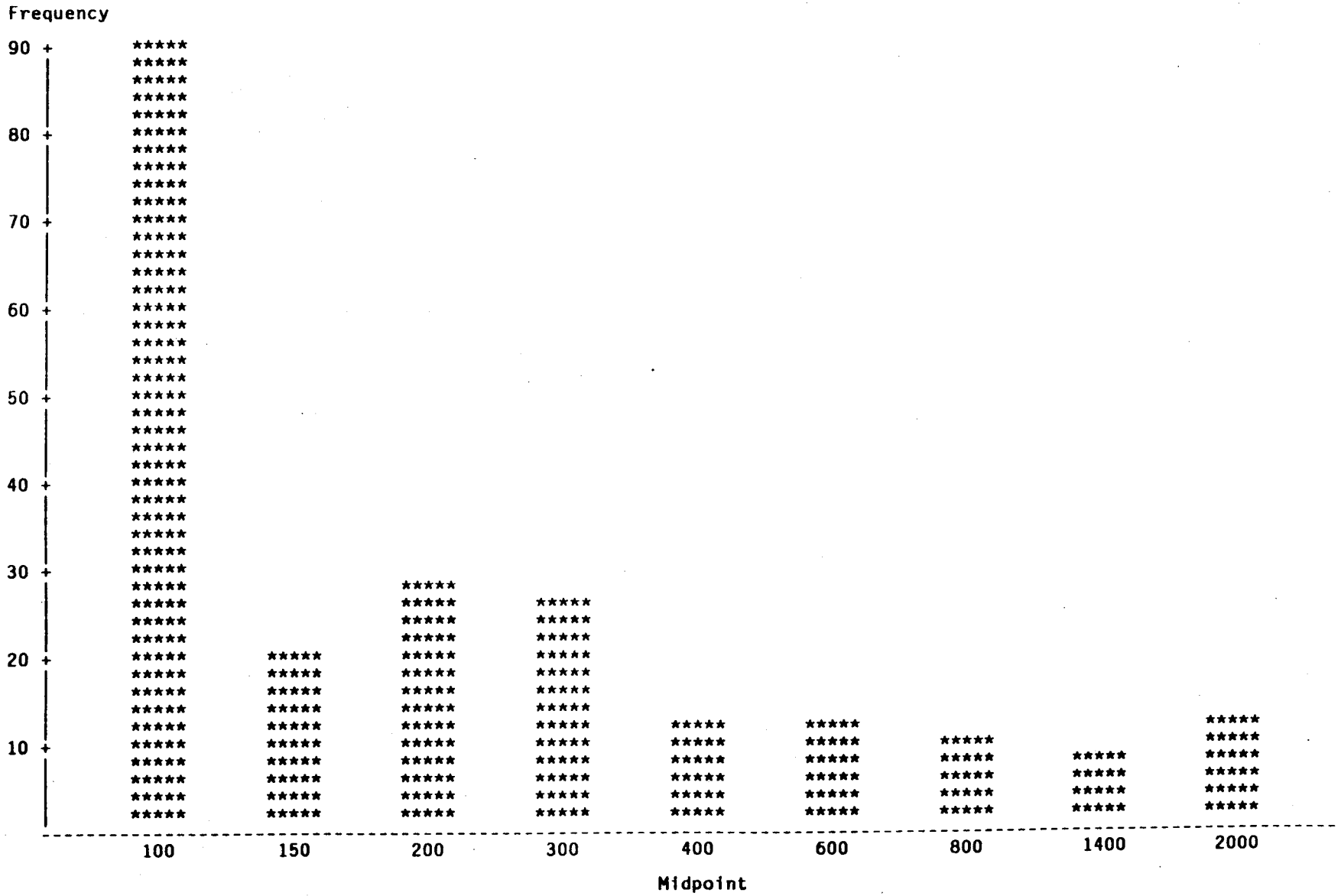

Frequency
45
40
35
30
25
20
15
10
5
4
12
20
28
36
44
52
60
68
76
84
92
100
108
116
124
132
140
148
156
Midpoint

(D)



(E)

Figure 3.7: Fig 3.7 Contd.


Frequency
90
80
70
60
50
40
30
20
10
100
150
200
300
400
600
800
1400
2000
Midpoint

(F)

Chapter 4

Performance of the Locally Optimum Detector for Multivariate Student-T and K-Distributed Disturbances

In radar problems involving weak signal applications, it is found that the large returns due to clutter can lead to a small signal to disturbance ratio. When the density function of the clutter exhibits an extended tail behavior relative to the Gaussian PDF, the probability density function of the disturbance can no longer be modeled as Gaussian. The significance of a non-Gaussian PDF with an extended tail is that many more large returns result than would be the case for a Gaussian PDF having the same variance. Hence, there is a need to be able to model non-Gaussian random processes.

The multivariate student-T distribution is a member of the class of joint PDFs arising from Spherically Invariant Random Processes (SIRP). SIRPs are explained in Chapter 2. When an SIRP is sampled at N instants in time, the resulting vector is said to be a spherically invariant random vector (SIRV). The theory of SIRPs offers a way to model the joint density function on these N samples where the correlation between the individual random variables in the vector is accounted for. With this approach locally optimum detector structures can be derived for non-Gaussian disturbances without the need to assume that the random variables are statistically independent. In this chapter we analyze the performance of the LOD for the known signal problem

when the background disturbance consisting of clutter and noise can be modeled as having a multivariate student-T distribution or a multivariate K-distribution.

4.1 The Multivariate Student-T Distribution

A convenient procedure for deriving the multivariate student-T distribution from the representation theorem [28] is discussed in this section. Let the random vector $\underline{X}$ have a multivariate Gaussian distribution with zero mean and covariance matrix M. The zero mean assumption will not affect the generality of the results that follow. The joint density function on the elements of $\underline{X}$ is given by

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^N |M|^{1/2}} e^{\frac{-\underline{x}^T M^{-1} \underline{x}}{2}} \tag{4.1}$$

where the vector $\underline{X}$ has $2N$ elements from N inphase and N quadrature samples. Consider the vector $\underline{W} = \underline{X}/\nu$, where ν is a nonnegative random variable statistically independent of $\underline{X}$. Let $\underline{w}^T M^{-1} \underline{w}$ be denoted by the variable p. Then, the conditional density function of the vector $\underline{W}$ given ν can be written as

$$f_{\underline{W}}(\underline{w}|\nu) = \frac{1}{(2\pi)^N |M|^{1/2}} \nu^{2N} e^{\frac{-\nu^2 p}{2}}. \tag{4.2}$$

The unconditional density function on $\underline{W}$ is given by

$$f_{\underline{W}}(\underline{w}) = \int_0^\infty f_{\underline{w}}(\underline{w}|\nu) f_\nu(\nu) d\nu \tag{4.3}$$

where $f_\nu(\nu)$ is the probability density function of the random variable ν. Because $\underline{X}$ and ν are statistically independent, it follows that

$$E(\underline{W}) = E(\frac{\underline{X}}{\nu}) = E(\underline{X}) E(\nu^{-1}) = \underline{0} \tag{4.4}$$

$$E(\underline{W}\,\underline{W}^T) = E(\underline{X}\,\underline{X}^T) E(\nu^{-2}) = E(\nu^{-2}) M. \tag{4.5}$$

It can be seen from the above equation that the covariances of the elements of the vector $\underline{W}$ can be adjusted by appropriate choice of $E(\nu^{-2})$.

With respect to equation (4.3), let $f_\nu(\nu)$ be the generalized chi PDF given by

$$f_\nu(\nu) = 2\frac{\nu^{2\beta-1} e^{-\alpha\nu^2} \alpha^\beta}{\Gamma(\beta)} \qquad \nu \geq 0. \tag{4.6}$$

From equation (4.6), $E(\nu^{-2})$ can be calculated. Specifically,

$$E(\nu^{-2}) = \int_0^\infty 2\nu^{-2}\frac{\nu^{2\beta-1}e^{-\alpha\nu^2}\alpha^\beta}{\Gamma(\beta)}d\nu = \int_0^\infty 2\frac{\nu^{2\beta-3}e^{-\alpha\nu^2}\alpha^\beta}{\Gamma(\beta)}d\nu. \tag{4.7}$$

Letting $\alpha\nu^2 = x$ in the above equation we get

$$E(\nu^{-2}) = \alpha\int_0^\infty \frac{x^{\beta-2}e^{-x}dx}{\Gamma(\beta)} = \alpha\frac{\Gamma(\beta-1)}{\Gamma(\beta)} = \frac{\alpha}{\beta-1}. \tag{4.8}$$

If we let $\alpha = \beta - 1$, then the generalized chi PDF in equation (4.6) is such that $E(\nu^{-2}) = 1$ irrespective of the choice for the parameter β. Then the generalized chi PDF takes the form

$$f_\nu(\nu) = \frac{2\nu^{2\beta-1}e^{-(\beta-1)\nu^2}(\beta-1)^\beta}{\Gamma(\beta)} \qquad \beta > 1. \tag{4.9}$$

In general, we can set the value of $E(\nu^{-2})$ to a desired constant c by choosing $\alpha = c(\beta-1)$. Then the covariance matrix of $\underline{W}$ is guaranteed to equal cM independent of β.

Integrating the conditional density function $f_{\underline{W}}(\underline{w}|\nu)$, as given by equation (4.2), over the PDF of the nonnegative random variable ν, we obtain the multivariate student-T distribution. The details are given below. Choosing $\alpha = \beta - 1$ in equation (4.6) we can write

$$\begin{aligned} f_{\underline{W}}(\underline{w}) &= \int_0^\infty \cdot \frac{1}{(2\pi)^N|M|^{1/2}}\nu^{2N}e^{\frac{-\nu^2 p}{2}}\frac{2\nu^{2\beta-1}e^{-(\beta-1)\nu^2}(\beta-1)^\beta}{\Gamma(\beta)}d\nu \\ &= \frac{(\beta-1)^\beta}{(2\pi)^N|M|^{1/2}\Gamma(\beta)}\int_0^\infty 2\nu^{2N+2\beta-1}e^{-\nu^2(\beta-1+p/2)}d\nu. \end{aligned} \tag{4.10}$$

Letting $(\beta - 1 + p/2)\nu^2 = y$ we get

$$\begin{aligned} f_{\underline{W}}(\underline{w}) &= \frac{(\beta-1)^\beta}{(2\pi)^N|M|^{1/2}\Gamma(\beta)}\int_0^\infty \frac{y^{N+\beta-1}e^{-y}}{(\beta-1+p/2)^{N+\beta}}dy \\ &= \frac{(\beta-1)^\beta\Gamma(N+\beta)}{(2\pi)^N|M|^{1/2}\Gamma(\beta)(\beta-1+p/2)^{N+\beta}}. \end{aligned} \tag{4.11}$$

The above expression is defined to be the $2N$-dimensional multivariate student-T distribution with parameters N and β. N represents the number of complex samples and β determines the tail behavior of the multivariate density function. The smaller

the value of β, the larger will be the tail.

The density function in equation (4.9) can be simulated as follows. The first step is to generate a standard Gamma variate having the density function $f_Y(y) = \frac{y^{\beta-1}e^{-y}}{\Gamma(\beta)}$. This was done using the subroutine DGAMDF from the IMSL package. The next step is to divide the generated random variable by the parameter $\beta-1$. Let $Z = Y/(\beta-1)$. The density function of Z is

$$f_Z(z) = \frac{(\beta-1)^\beta z^{\beta-1}e^{-z(\beta-1)}}{\Gamma(\beta)}. \tag{4.12}$$

The positive square root of $\frac{Y}{\beta-1}$ results in the desired density function. Let $\nu = Z^{\frac{1}{2}}$. Therefore $Z = \nu^2$. Introducing the Jacobian of the transformation, the density function of ν becomes

$$f_\nu(\nu) = \frac{2\nu^{2\beta-1}e^{-(\beta-1)\nu^2}(\beta-1)^\beta}{\Gamma(\beta)} \tag{4.13}$$

which is identical to that in equation (4.9).

4.1.1 The Locally Optimum Detector

The locally optimum detector for the multivariate student-T distribution can now be derived. From equation (2.32) the locally optimum detector is given as

$$\frac{\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0}}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{4.14}$$

Assuming the disturbance can be modeled by a multivariate student-T distribution, $f_{\underline{D}}(\underline{r})$ is given by equation (4.11), with the variable $\underline{R}$ replacing the variable $\underline{W}$, where $p = \underline{r}^T M^{-1}\underline{r}$. Since equation (4.14) is a ratio test and all constants can be placed in the threshold which is determined by specifying a false alarm probability, all multiplicative constants are ignored for convenience. Hence, we will be concerned only with the terms containing the variable $\underline{R}$. Excluding the constant term the numerator in the ratio test is given by

$$\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0} = \frac{\partial}{\partial\theta}\left[\frac{1}{(\beta-1+p/2)^{N+\beta}}\right]|_{\theta=0}. \tag{4.15}$$

Applying the chain rule, the derivative with respect to θ can be expressed as the derivative with respect to p times the derivative of p with respect to θ. The derivative

of p with respect to θ at $\theta = 0$ can be derived as

$$\frac{\partial p}{\partial \theta}|_{\theta=0} = (\frac{\partial}{\partial \theta}(\underline{r} - \theta \underline{s})^T M^{-1}(\underline{r} - \theta \underline{s}))|_{\theta=0} = -2\underline{s}^T M^{-1}\underline{r}. \tag{4.16}$$

Therefore, the numerator in the ratio test, excluding the constant, is given by

$$\frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta}|_{\theta=0} = (\beta - 1 + p/2)^{-(N+\beta+1)} \times \underline{s}^T M^{-1}\underline{r}. \tag{4.17}$$

From the above equation, the sufficient statistic for the locally optimum detector for the multivariate student-T distribution can be written as

$$T_{LOD}(\underline{r}) = \frac{\underline{s}^T M^{-1}\underline{r}}{\beta - 1 + p/2}. \tag{4.18}$$

The above result for the LOD statistic is very significant. The numerator in equation (4.18) is recognized as the Gaussian linear detector. This detector is a matched filter which maximizes the signal-to-disturbance ratio whether or not the disturbance is Gaussian. In weak signal applications, by definition, the signal to disturbance ratio will still be low after matched filtering. The denominator of the LOD statistic is the nonlinear term in the statistic. The behavior of the nonlinearity is such that it scales down large values of p and enhances small values of p. The nonlinearity is plotted as a function of p in Fig. 4.1. This is reasonable because, as shown in Section 1.3, large values of radar returns result in large p while small values of the returns yields small values of p. Because it is known a priori that we are dealing with the weak signal problem, large returns cannot be due to the signal. Consequently, the output of the matched filter is weighted by a small number. On the other hand, the matched filter output is weighted by a large number when the return is small and the contribution due to the signal, if present, can be detected. However, when the signal is present, the optimum nonlinearity alone is not sufficient to get detections, though it brings the output close to the threshold value. The matched filter which has a higher output value when the signal is present than when the signal is absent, contributes to raising the output value over the threshold when detections are obtained. The role of the matched filter is explained in greater detail in Section 4.1.3.

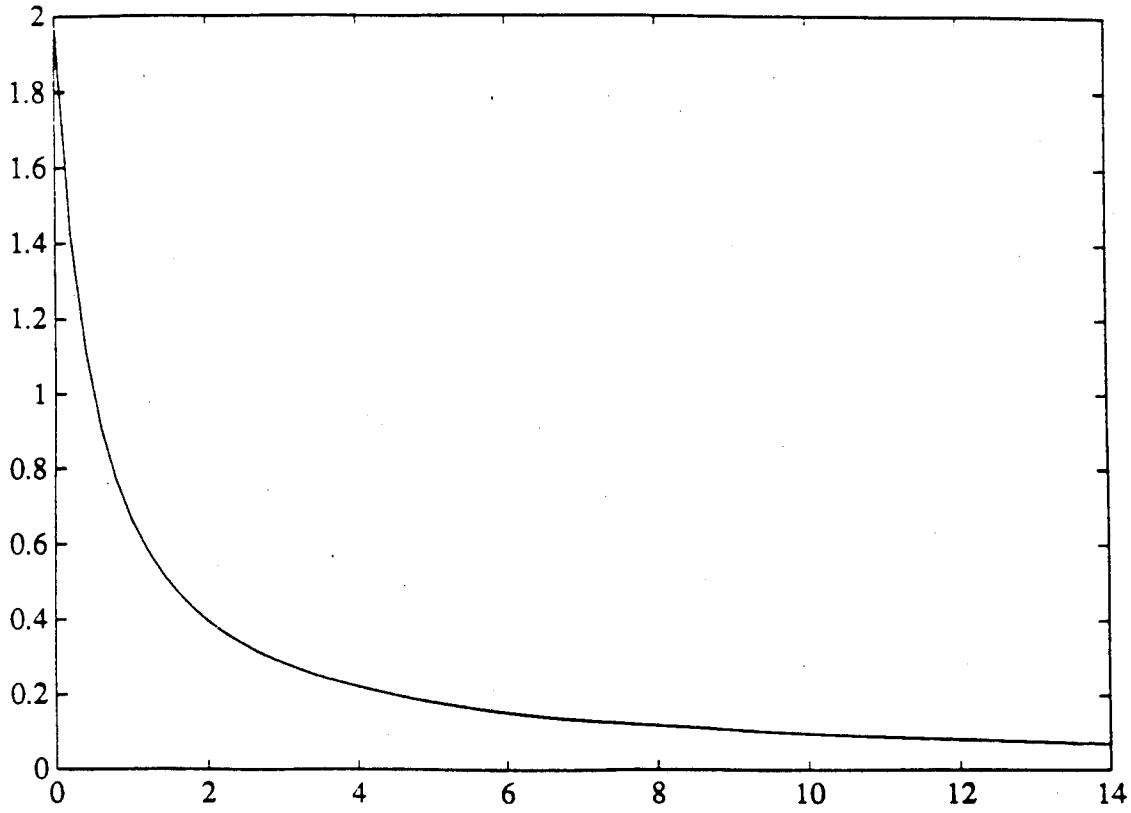


Figure 4.1: Nonlinearity of the LOD statistic for the student-T distribution

4.1.2 Computer Simulation of Performance

Because analytical closed form expressions for the detection and false alarm probabilities of the locally optimum detector in a multivariate student-T distributed clutter are difficult to obtain, performance is evaluated through computer simulations for weak signal applications. For simulation purposes a multivariate student-T distributed disturbance vector $\underline{D}$ and a transmitted signal vector $\underline{S}$ have to be generated. The covariance matrix of the clutter process is assumed known with unit elements along the diagonal. To get the covariance matrix of the disturbance we add a small number, determined by the clutter to noise ratio, to the diagonal elements of the clutter covariance matrix. This serves to limit the performance of the receiver even where the clutter power is negligible. In this simulation, the clutter to noise ratio is taken to be 80 dB. The simulation procedure for the disturbance vector is outlined as follows:

1. Generate a $2N$-dimensional white Gaussian random vector $\underline{X}'$. This was done by using the DRRNOA subroutine from the IMSL package.

2. Do a Cholesky decomposition on the matrix M to get $M = KK^T$ where K is a lower triangular matrix.

3. The vector $\underline{X} = K\underline{X}'$ is the multivariate correlated Gaussian vector.

4. Generate a standard Gamma variate Y. This was done by using the subroutine DGAMDF from the IMSL package.

5. Obtain $\nu = (\frac{Y}{\beta-1})^{\frac{1}{2}}$. The random variable ν has the generalized chi PDF.

6. Obtain the multivariate student-T distributed disturbance vector $\underline{D}$ with the desired correlation properties from $\underline{D} = \frac{\underline{X}}{\nu}$.

The block diagram of the simulation procedure is shown in Fig. 4.2. The autocorrelation of the clutter process is taken to be a geometric function in this problem. Assuming radar returns from clutter cells to be highly correlated, as is the case with ground clutter, the sample to sample time correlation is taken as 0.95 in this problem. Specifically, the sample autocorrelation function is chosen as

$$R_{CC}(n) = (0.95)^n \quad n = 0, 1, ..., N-1 \tag{4.19}$$

where $R_{CC}(n)$ is the discrete time autocorrelation function of the clutter process.

Figures 4.3 and 4.4 show the autocorrelation and the power spectral density of the clutter process. The power spectral density of clutter has a very small spread as the clutter is highly correlated. Using equation (4.19) the elements of the covariance matrix of the disturbance can be filled appropriately. The elements of the signal vector are chosen such that the n^{th} element $S_n = e^{j2\pi f_D(n-1)T}$, $n = 1, 2, ..., N$. f_D represents the Doppler frequency shift of the received signal and T represents the time separation between sampling instants.

The detector in equation (4.18) is now simulated. A value of $\beta = 1.5$ for the multivariate student-T distribution is chosen because this value results in a relatively long tail for the corresponding marginal PDF of one element of the vector. By evaluating thresholds for specified false alarm probabilities, the student-T distribution was seen to have heavier tails than the Gaussian distribution for false alarm probabilities less than 10^{-4} but smaller tails than the Gaussian otherwise.

The thresholds corresponding to false alarm probabilities 10^{-k}; $k = 1, 2, 3, 4$ are obtained through the method of extreme value theory explained in Chapter 3. Once the threshold is set the detection probabilities are obtained by simulating the LOD for received vectors consisting of the sum of the signal and disturbance vectors for various signal-to-disturbance ratios. The value of f_D is chosen to be zero in this simulation, resulting in a worst case situation. The number of trials in the Monte Carlo simulation for obtaining detection probabilities is equal to $10,000$. The performance of the LOD is compared to that of the Gaussian detector for the same multivariate student-T distributed clutter. The test statistic for the Gaussian detector is the same as the numerator of the LOD, which is $\underline{s}^T M^{-1} \underline{r}$.

4.1.3 Results of the Computer Simulation

The results of the computer simulations are shown in Tables 4.1-4.10. When SCR is less than 0 dB, it is seen from the tables that the LOD always outperforms the Gaussian receiver for all values of the tabulated false alarm probabilities. The difference is especially significant for false alarm probabilities equal to 10^{-3} and 10^{-4}.

The student-T distribution, while being heavier tailed than the Gaussian, is not as heavy tailed as the K-distribution and Weibull distribution. In fact, the student-T distribution may not be a likely candidate for modeling the radar disturbance. The student-T distribution was chosen as the first distribution to be studied only because

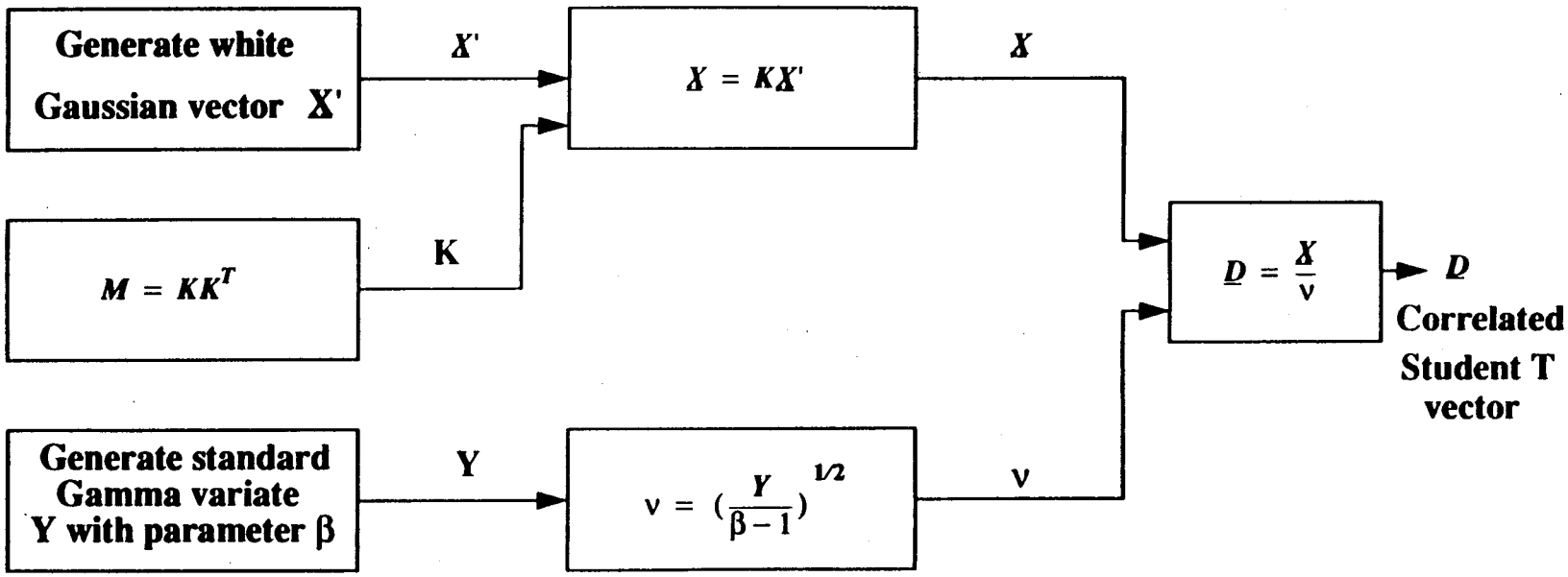


Figure 4.2: Generation scheme for the correlated multivariate student-T distributed vector

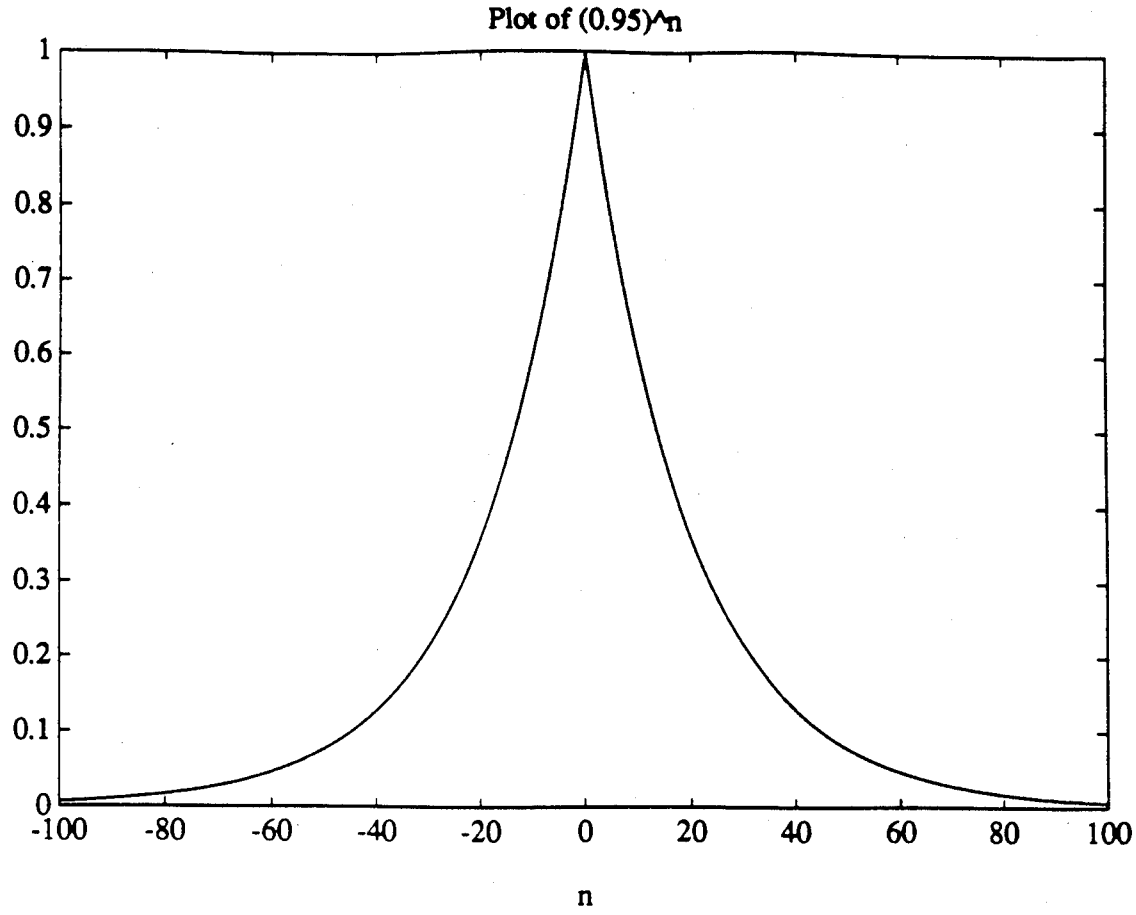


Figure 4.3: Autocorrelation of the clutter process

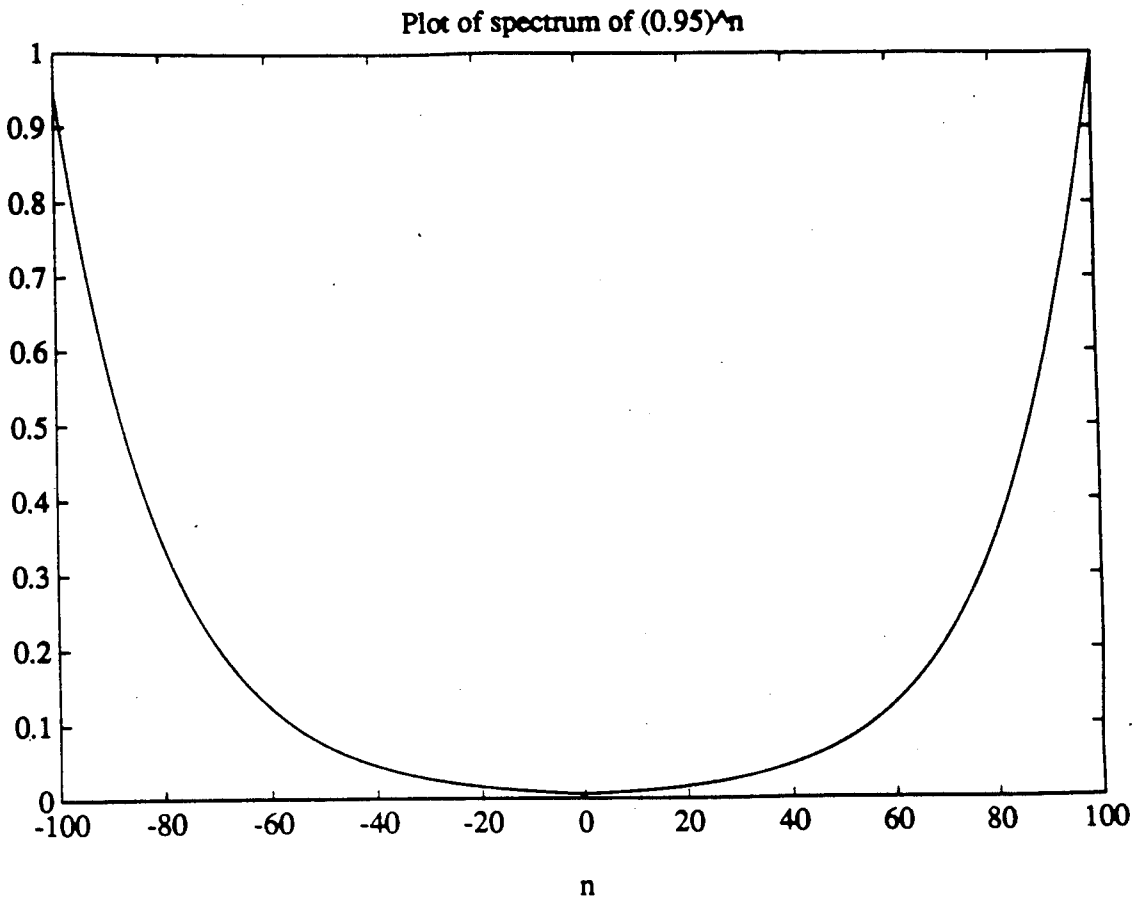


Figure 4.4: Power spectral density of the clutter process

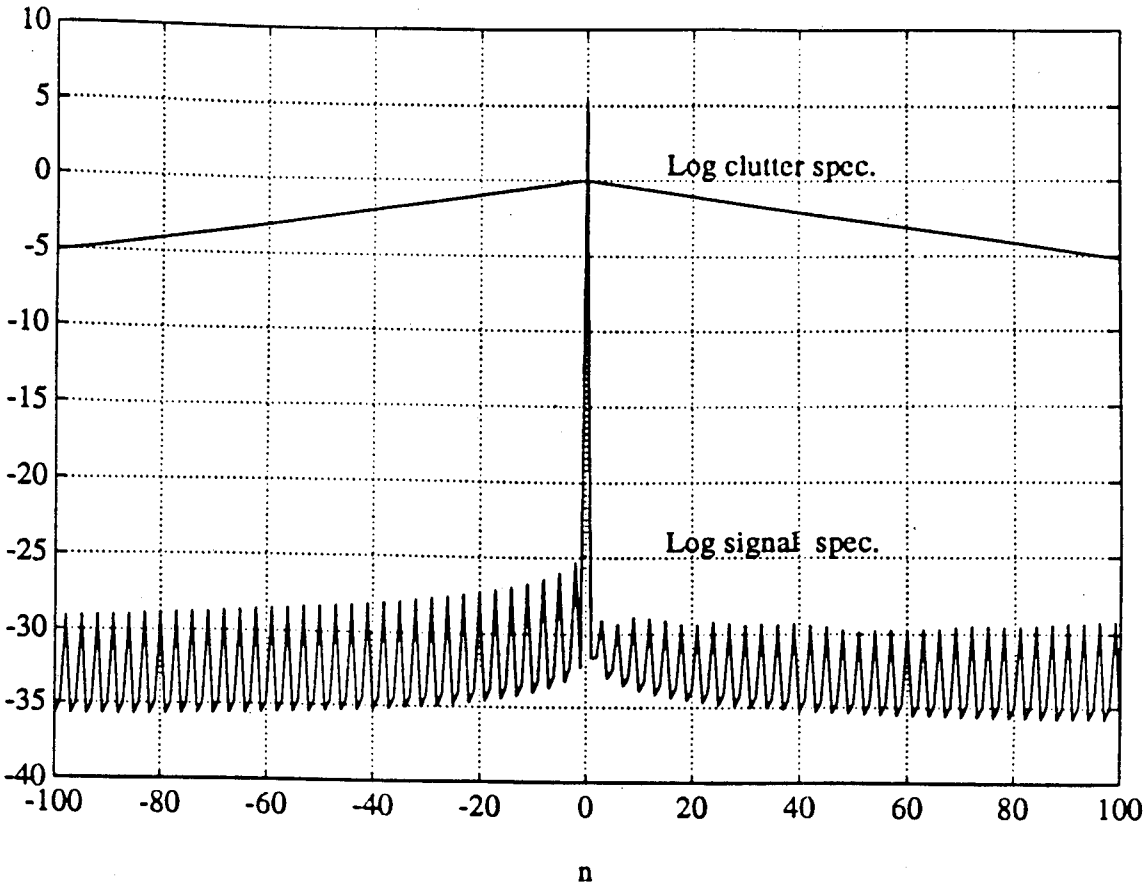


Figure 4.5: Log power spectral density of the clutter and signal processes

of the mathematical simplicity and well behaved nature of its multivariate PDF. Nevertheless, the analysis done with the student-T distribution confirms that the LOD outperforms the Gaussian receiver for a non-Gaussian weak signal application.

As the signal to clutter ratio is reduced, it can be observed from the tables that the Gaussian receiver performance degrades abruptly for false alarm probabilities less than or equal to 10^{-2} whereas the LOD shows a gentler variation in performance. From the Tables it is seen that the performance of the LOD peaks around 0 dB and falls off for both higher and lower values of SCR. But as the SCR values fall below 0 dB, the degradation in the LOD's performance is not as drastic as that of the Gaussian receiver. Both the receivers show an improvement in performance as the number of samples is increased. However, the LOD shows a dramatic improvement in performance when the sample size is greater than 64. When the sample size is equal to 64 and the false alarm probability is as low as 10^{-4}, it is seen from Table 4.7 that the detection probability resulting from the LOD is on the order of tenths for SCR as small as -8 dB. The corresponding detection probabilities resulting from the Gaussian receiver are negligibly small. In fact, there is an improvement factor in the vicinity of three orders of magnitude in favor of the LOD. When -10 dB$\leq$SCR$<$0 dB, an interesting observation from the tables, for all values of N considered is that the LOD outperforms the Gaussian receiver by close to one order of magnitude for $P_F = 10^{-2}$, by two orders of magnitude for $P_F = 10^{-3}$ and by three orders of magnitude for $P_F = 10^{-4}$. For SCR values lower than -10 dB, the LOD significantly outperforms the Gaussian receiver but with very small values of P_D. When both SCR$\geq$4 dB and $P_F > 10^{-3}$, the Gaussian receiver outperforms the LOD, as can be seen from the tables. For positive values of SCR in the range 0-5 dB and for false alarm probabilities equal to 10^{-3} and 10^{-4}, it is interesting to note that the LOD still outperforms the Gaussian receiver. The LOD shows a significant performance improvement over the Gaussian receiver over a dynamic range of about 14 dB. The end points of the range vary depending on the sample size and false alarm probability.

The test statistic is a product involving the outputs of matched filter and the optimum nonlinearity. In section 4.1.1 it was explained how the nonlinearity present in the test statistic boosts small values of the received signal and attenuates large values. This not only serves to bring down the value of the threshold needed to

obtain the desired false alarm probability but also to bring the output for small received signals close to the threshold value whether or not the desired signal is present. The role of the matched filter is explained as follows. The matched filter has a larger output value when the signal is present as opposed to the signal absent case. This serves to increase the statistic in equation (4.18). However, the quadratic form p, in general, also has a larger value when the signal is present than under the H_0 hypothesis. Thus, the factor in the test statistic in equation (4.18) due to the nonlinearity decreases when the signal is present. This serves to lower the value of the test statistic. Therefore, detections are obtained only when the increase in the due to the matched filter dominates the decrease due to the nonlinearity. Simulations reveal that the matched filter's role is dominant only when the signal to clutter ratio is in the range -10 dB$\leq$SCR$<$0 dB. This is expected since the linear receiver's performance is known to drop drastically for very low signal to clutter ratios. The matched filter's effect is enhanced when a Doppler is present in the desired received signal since the clutter components become less correlated with the Doppler shifted reference signal. However, one must be careful when the Doppler frequency is so large that the signal spectrum appears in the tail of the clutter spectrum. Then a strong signal situation may exist and the nonlinearity which transforms large values into small values will cause performance to degrade. The LOD should not be used in strong signal situations.

The aim of using a LOD is to obtain detection in Range-Doppler-Azimuth cells where conventional space-time processing is unable to obtain acceptable performance. In present day radars these cells are ignored because the probability of detection is so small under a false alarm constraint. In general, when the SCR is relatively high ($>$ 0 dB) the likelihood ratio test is the optimal test for target detection under a fixed false alarm constraint. In addition to not performing well when SCR is too large, the LOD does not perform well when SCR is too small. When the signal to clutter ratio drops below a certain value, depending upon N and P_F, the LOD receiver hardly shows any detections even though it still outperforms the Gaussian receiver. This is because the PDFs under H_0 and H_1 are so close to each other that it is impossible to discriminate between them without increasing the sample size by orders of magnitude.

The concepts of spherically invariant random processes and locally optimum de-

SCR		LOD	GR
5 dB	P_D	0.35	0.65
3 dB	P_D	0.36	0.35
1 dB	P_D	0.37	0.23
0 dB	P_D	0.38	0.11
-3 dB	P_D	0.32	0.052
-6 dB	P_D	0.27	0.039
-8 dB	P_D	0.23	0.03
-9 dB	P_D	0.15	0.024
-10 dB	P_D	0.10	0.019

Table 4.1: N=16, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
5 dB	P_D	0.08	0.007
3 dB	P_D	0.11	0.005
1 dB	P_D	0.17	0.004
0 dB	P_D	0.16	0.003
-3 dB	P_D	0.13	0.002
-6 dB	P_D	0.10	0.001
-8 dB	P_D	0.08	0.001

Table 4.2: N=16, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

tectors are particularly relevant in the context of modern radar applications. When the radar scans a volume searching for targets there might be certain regions in the volume where the clutter returns are significantly stronger than the desired target returns. It is in these regions that we can obtain detections with the LODs. There is a need to monitor the environment so that we are able to separate the clutter regions from volumes that are limited by weak background noise. When detection is limited by background noise alone, the LOD is not applicable. Using the concepts of artificial intelligence, clutter patches can be identified and the underlying multivariate PDF of the clutter returns can be approximated using the library of SIRPs that have been developed [26]. From the library of LODs the LOD corresponding to the approximated SIRP can be used in clutter regions to obtain detections if the target is present, where earlier it would not have been possible.

SCR		LOD	GR
5 dB	P_D	0.38	0.99
3 dB	P_D	0.42	0.57
1 dB	P_D	0.45	0.32
0 dB	P_D	0.46	0.18
-3 dB	P_D	0.38	0.08
-6 dB	P_D	0.31	0.06
-8 dB	P_D	0.27	0.04
-9 dB	P_D	0.20	0.033
-10 dB	P_D	0.13	0.026
-11 dB	P_D	0.10	0.019

Table 4.3: N=32, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
8 dB	P_D	0.09	0.15
7 dB	P_D	0.11	0.08
5 dB	P_D	0.14	0.014
2 dB	P_D	0.22	0.008
0 dB	P_D	0.26	0.004
-3 dB	P_D	0.19	0.003
-6 dB	P_D	0.14	0.002
-8 dB	P_D	0.11	0.002
-9 dB	P_D	0.08	0.001

Table 4.4: N=32, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
5 dB	P_D	0.44	0.96
3 dB	P_D	0.47	0.68
1 dB	P_D	0.53	0.42
0 dB	P_D	0.55	0.30
-3 dB	P_D	0.48	0.17
-6 dB	P_D	0.40	0.10
-8 dB	P_D	0.36	0.03
-10 dB	P_D	0.14	0.023
-12 dB	P_D	0.09	0.012

Table 4.5: N=64, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
9 db	P_D	0.10	0.17
7 dB	P_D	0.15	0.10
5 dB	P_D	0.18	0.02
0 dB	P_D	0.36	0.005
-3 dB	P_D	0.28	0.002
-6 dB	P_D	0.22	0.002
-8 dB	P_D	0.18	0.001
-9 dB	P_D	0.10	0.001

Table 4.6: N=64, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
5 dB	P_D	0.08	0.0007
0 dB	P_D	0.25	0.0002
-3 dB	P_D	0.19	0.0002
-6 dB	P_D	0.13	0.0001
-8 dB	P_D	0.11	0.0001
-9 dB	P_D	0.06	0.0001

Table 4.7: N=64, $P_F = 10^{-4}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

4.2 The Multivariate K-Distribution

In the previous section we analyzed the performance of the LOD for the multivariate student-T distribution. The multivariate K-distribution is also a member of the class of joint PDFs arising from SIRPs. Jakeman [50] has shown that the K-distributed PDF has a physical interpretation in the sense that it arises from the random walk problem where the number of steps itself is random having a negative binomial distribution. Also, radar clutter has been empirically shown to have K-distributed PDF. In this chapter we analyze the performance of the LOD when the background disturbance consisting of the clutter and noise can be approximated as having a multivariate K-distribution.

Derivation of the multivariate K-distributed PDF from the representation theorem for SIRPs [28] is discussed next. Let the random vector $\underline{X}$ have a multivariate Gaussian distribution with zero mean and covariance matrix M. The zero mean assumption will not affect the generality of the results that follow. The joint density function on the elements of $\underline{X}$ is given by equation (4.1). Consider the vector

SCR		LOD	GR
5 dB	P_D	0.51	0.998
0 dB	P_D	0.65	0.71
-1 dB	P_D	0.63	0.57
-3 dB	P_D	0.57	0.35
-6 dB	P_D	0.51	0.16
-8 dB	P_D	0.46	0.11
-10 dB	P_D	0.22	0.032
-12 dB	P_D	0.15	0.027
-13 dB	P_D	0.11	0.023
-15 dB	P_D	0.09	0.02

Table 4.8: N=128, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
5 dB	P_D	0.20	0.30
4 dB	P_D	0.23	0.23
3 dB	P_D	0.29	0.19
0 dB	P_D	0.48	0.01
-3 dB	P_D	0.36	0.006
-6 dB	P_D	0.30	0.005
-8 dB	P_D	0.26	0.004
-9 dB	P_D	0.14	0.003
-10 dB	P_D	0.09	0.002

Table 4.9: N=128, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
7 dB	P_D	0.08	0.001
5 dB	P_D	0.13	0.0.0004
2 db	P_D	0.30	0.0004
0 dB	P_D	0.37	0.0003
-3 dB	P_D	0.27	0.0003
-6 dB	P_D	0.21	0.0002
-8 dB	P_D	0.17	0.0001
-9 dB	P_D	0.09	0.0001

Table 4.10: N=128, $P_F = 10^{-4}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

$\underline{W} = \nu \underline{X}$, where ν is a nonnegative random variable statistically independent of $\underline{X}$. Let $\underline{w}^T M^{-1} \underline{w}$ be once again denoted by the variable p. Then, the conditional density function of the vector $\underline{W}$ given ν can be written as

$$f_{\underline{W}}(\underline{w}|\nu) = \frac{1}{(2\pi)^N |M|^{1/2}} \nu^{-2N} e^{-\frac{p}{2\nu^2}}. \tag{4.20}$$

The unconditional density function on $\underline{W}$ is given by equation (4.3), where $f_\nu(\nu)$ is the probability density function of the random variable ν. Because $\underline{X}$ and ν are statistically independent, it follows that

$$E(\underline{W}) = E(\nu \underline{X}) = E(\underline{X})E(\nu) = \underline{0} \tag{4.21}$$

$$E(\underline{W}\,\underline{W}^T) = E(\underline{X}\,\underline{X}^T)E(\nu^2) = E(\nu^2)M. \tag{4.22}$$

As is the case for the student-T distribution, the variance of the elements of the vector $\underline{W}$ can be adjusted by appropriate choice of $E(\nu^2)$. Let $f_\nu(\nu)$ be the generalized chi PDF given by equation (4.6). $E(\nu^2)$ is then given by

$$E(\nu^2) = \int_0^\infty 2\nu^2 \frac{\nu^{2\beta-1} e^{-\alpha\nu^2} \alpha^\beta}{\Gamma(\beta)} d\nu = \int_0^\infty 2 \frac{\nu^{2\beta+1} e^{-\alpha\nu^2} \alpha^\beta}{\Gamma(\beta)} d\nu \tag{4.23}$$

Letting $\alpha\nu^2 = x$ in the above equation we get

$$E(\nu^2) = \int_0^\infty \frac{x^\beta e^{-x} dx}{\alpha \Gamma(\beta)} = \frac{\Gamma(\beta+1)}{\alpha\Gamma(\beta)} = \frac{\beta}{\alpha}. \tag{4.24}$$

If we let $\alpha = \beta$, then the generalized chi PDF in equation (4.6) is such that $E(\nu^2) = 1$ irrespective of the choice for the parameter β. Then the generalized chi PDF takes the form

$$f_\nu(\nu) = \frac{2\nu^{2\beta-1} e^{-\beta\nu^2} \beta^\beta}{\Gamma(\beta)} \quad \beta > 1. \tag{4.25}$$

In general, we can set the value of $E(\nu^2)$ to a desired constant C by choosing α appropriately. Integrating the conditional density function $f_{\underline{W}}(\underline{w}|\nu)$ as given by equation (4.20), over the PDF of the nonnegative random variable ν, we obtain the multivariate K-distribution. The details are given below. Choosing $\alpha = \beta$ in equation (4.6)

we can write

$$f_{\underline{W}}(\underline{w}) = \int_0^\infty \frac{1}{(2\pi)^N |M|^{1/2}} \nu^{-2N} e^{-\frac{p}{2\nu^2}} \frac{2\nu^{2\beta-1} e^{-\beta\nu^2} \beta^\beta}{\Gamma(\beta)} d\nu$$
$$= \frac{\beta^\beta}{(2\pi)^N |M|^{1/2} \Gamma(\beta)} \int_0^\infty 2\nu^{-2N+2\beta-1} e^{-\beta\nu^2 - \frac{p}{2\nu^2}} d\nu. \quad (4.26)$$

Letting $\beta\nu^2 = y$ and simplifying we get

$$f_{\underline{W}}(\underline{w}) = \frac{\beta^N}{(2\pi)^N |M|^{1/2} \Gamma(\beta)} \int_0^\infty y^{-N+\beta-1} e^{-(y+\frac{\beta p}{2y})} dy. \quad (4.27)$$

From page 183 of Watson's book on Bessel functions [51], we have the result

$$K_\beta(z) = \frac{1}{2}(\frac{z}{2})^\beta \int_0^\infty y^{-\beta-1} e^{-(y+\frac{z^2}{4y})} dy \quad (4.28)$$

provided that the real part of $z^2 > 0$. $K_\beta(z)$ represents the modified Bessel function of the second kind of order β. Combining equations (4.27) and (4.28) results in

$$f_{\underline{W}}(\underline{w}) = 2\frac{\beta^N}{(2\pi)^N |M|^{1/2} \Gamma(\beta)} (\frac{2}{(2\beta p)^{1/2}})^{N-\beta} K_{N-\beta}[(2\beta p)^{1/2}]$$
$$= \frac{2^{-\frac{N+\beta}{2}+1} \beta^{\frac{N+\beta}{2}} K_{N-\beta}[(2\beta p)^{1/2}]}{\pi^N |M|^{1/2} \Gamma(\beta) p^{\frac{N-\beta}{2}}}. \quad (4.29)$$

The above expression is defined to be the $2N$-dimensional multivariate K-distribution with parameters N and β. N represents the number of complex samples and β determines the tail behavior of the multivariate density function.

For simulation purposes, the density function in equation (4.25) can be simulated as follows. The first step is to generate a standard Gamma variate having the density function $f_Y(y) = \frac{y^{\beta-1} e^{-y}}{\Gamma(\beta)}$. The IMSL package is used to generate the standard Gamma variates. The next step is to divide the generated random variable by the parameter β. Let $X = Y/\beta$. The density function of X is

$$f_X(x) = \frac{\beta^\beta x^{\beta-1} e^{-\beta x}}{\Gamma(\beta)}. \quad (4.30)$$

The positive square root of $\frac{Y}{\beta}$ results in a variate having the desired density function. Let $\nu = X^{\frac{1}{2}}$. Therefore, $X = \nu^2$. Introducing the Jacobian of the transformation,

the density function of ν becomes

$$f_\nu(\nu) = \frac{2\nu^{2\beta-1}e^{-\beta\nu^2}\beta^\beta}{\Gamma(\beta)} \tag{4.31}$$

which is identical to that in equation (4.25). This random variable is used to multiply the Gaussian vector $\underline{X}$ in order to generate the K-distributed vector $\underline{W}$.

4.2.1 The Locally Optimum Detector

The locally optimum detector for the multivariate K-distribution can now be derived. From equation (2.32) the locally optimum detector is given as

$$\frac{\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0}}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{4.32}$$

Assuming the disturbance can be modeled by a multivariate K-distribution, $f_{\underline{D}}(\underline{r})$ is given by equation (4.29), with the variable $\underline{R}$ replacing the variable $\underline{W}$, where $p = \underline{r}^T M^{-1}\underline{r}$. Since equation (4.32) is a ratio test and all constants can be placed in the threshold which is determined by specifying a false alarm probability, all multiplicative constants are ignored for convenience. Hence, we will be concerned only with the terms containing the vector $\underline{R}$. Excluding the constant term the numerator in the ratio test is given by

$$\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0} = \frac{\partial}{\partial\theta}[p^{\frac{\beta-N}{2}}K_{N-\beta}[(2\beta p)^{\frac{1}{2}}]|_{\theta=0}. \tag{4.33}$$

Applying the chain rule, the derivative with respect to θ can be expressed as the derivative with respect to p times the derivative of p with respect to θ. From equation 4.16, the derivative of p with respect to θ at $\theta = 0$ is given by $-2\underline{s}^T M^{-1}\underline{r}$. Therefore, the numerator in the ratio test, excluding the constant, is given by

$$\begin{aligned}\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}|_{\theta=0} &= (\frac{\beta-N}{2}p^{\frac{\beta-N}{2}-1}K_{N-\beta}[(2\beta p)^{\frac{1}{2}}] \\ &+ p^{\frac{\beta-N}{2}}K'_{N-\beta}[(2\beta p)^{\frac{1}{2}}]\frac{\beta^{\frac{1}{2}}}{2p}) \times -2\underline{s}^T M^{-1}\underline{r}\end{aligned} \tag{4.34}$$

where $K'_{N-\beta}[x]$ denotes the derivative of $K_{N-\beta}[x]$ with respect to x. From the above equation, the sufficient statistic for the locally optimum detector for the multivariate

K-distribution can be written as

$$T_{LOD}(\underline{r}) = -\underline{s}^T M^{-1}\underline{r}[\frac{\beta - N}{p} + (\frac{2\beta}{p})^{1/2}\frac{K'_{N-\beta}[(2\beta p)^{\frac{1}{2}}]}{K_{N-\beta}[(2\beta p)^{\frac{1}{2}}]}. \tag{4.35}$$

Equation (4.35) can be simplified further using the property of Bessel functions. From page 79 of Watson's book on Bessel functions [51] we have the two results:

$$1.\quad K'_\nu(z) = \tfrac{\nu}{z}K_\nu(z) - K_{\nu+1}(z). \tag{4.36}$$

$$2.\qquad K_{-\nu}(z) = K_\nu(z). \tag{4.37}$$

Use of equations (4.36-4.37) in equation (4.35) and combining the constant term with the threshold results in

$$T_{LOD}(\underline{r}) = \underline{s}^T M^{-1}\underline{r}(\frac{K_{N-\beta+1}[(2\beta p)^{\frac{1}{2}}]}{p^{\frac{1}{2}}K_{N-\beta}[(2\beta p)^{\frac{1}{2}}]}). \tag{4.38}$$

The numerator in equation (4.38) once again has a factor a term which is the Gaussian linear detector or the matched filter that maximizes the signal-to-disturbance ratio whether or not the disturbance is Gaussian. The factor multiplying the linear detector is the optimum nonlinearity for weak signal detection when the disturbance is K-distributed. The nonlinearity is plotted as a function of p in Fig. 4.5. The behavior of the nonlinearity is similar in form to the one obtained for the student-T distributed disturbance. It scales down large values of p and enhances small values of p. Since it is known that we are dealing with the weak signal problem, large returns cannot be due to the signal. Consequently, the output of the matched filter is weighted by a small number. On the other hand, when the return is small, there is a greater chance of the signal being detected, if present. Hence, the nonlinearity weights it by a large number. The role of the matched filter is also similar to the situation when the disturbance was modeled as having a student-T distributed disturbance.

4.2.2 Computer Simulation of Performance

Because analytical closed form expressions for the detection and false alarm probabilities of the locally optimum detector in a multivariate K-distributed clutter are difficult to obtain, performance is evaluated through computer simulations for weak signal applications. For simulation purposes a multivariate K-distributed disturbance vector $\underline{D}$ and a transmitted signal vector $\underline{S}$ have to be generated. The covariance

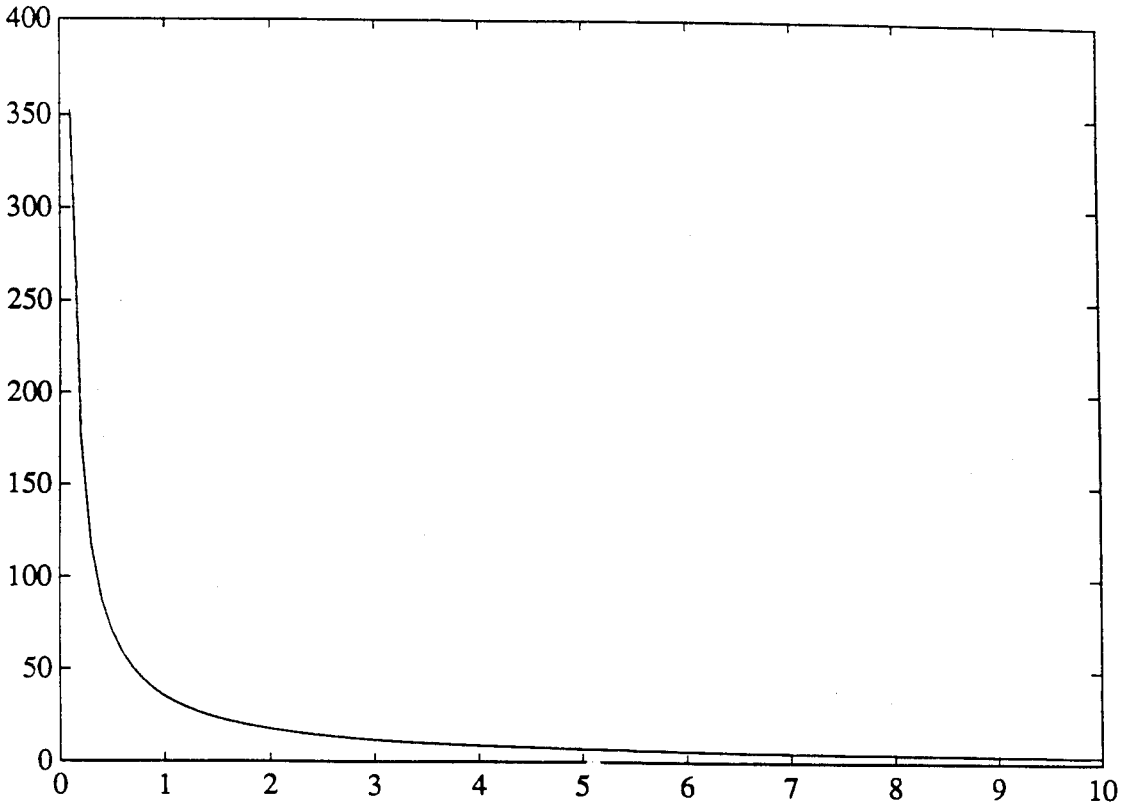


Figure 4.5: Nonlinearity for the K-distribution

matrix of the clutter process is assumed known with unit elements along the diagonal. To get M, the covariance matrix of the disturbance, we add a small number, determined by the clutter to noise ratio, to the diagonal elements of the clutter covariance matrix. This serves to limit the performance of the receiver even where the clutter power is negligible. In this simulation, the clutter to noise ratio is taken to be 80 dB. The simulation procedure for the disturbance vector is outlined as follows:

1. Generate a $2N$-dimensional white Gaussian random vector $\underline{X}'$. This was done by using the DRRNOA subroutine from the IMSL package.

2. Do a Cholesky decomposition on the matrix M to get $M = KK^T$ where K is a lower triangular matrix.

3. The vector $\underline{X} = K\underline{X}'$ is the multivariate correlated Gaussian vector.

4. Generate a standard Gamma variate Y. This was done by using the subroutine DGAMDF from the IMSL package.

5. Obtain $\nu = (\frac{Y}{\beta})^{\frac{1}{2}}$. The random variable ν has the generalized chi PDF.

6. Obtain the multivariate K-distributed disturbance vector $\underline{D}$ with the desired correlation properties from $\underline{D} = \nu\underline{X}$.

The block diagram of the simulation procedure is shown in Fig. 4.6. The autocorrelation of the clutter process is taken to be a Gaussian function in this problem. Assuming radar returns from clutter cells to be highly correlated, as is the case with ground clutter, the sample to sample time correlation is taken as 0.999 in this problem. Specifically, the sample autocorrelation function is chosen as

$$R_{CC}(n) = exp(-0.000801n^2) \quad n = 0, 1, ..., N-1 \tag{4.39}$$

where $R_{CC}(n)$ is the discrete time autocorrelation function of the clutter process. The autocorrelation and the power spectral density of the clutter process are shown in Figures 4.7 and 4.8 respectively. Once again, we see that the spectral spread of the clutter is very small. Using the above function the elements of the covariance matrix of the disturbance can be filled appropriately. The elements of the signal vector are chosen such that the n^{th} element $S_n = e^{j2\pi f_D(n-1)T}$, $n = 1, 2, ..., N$. f_D represents the Doppler frequency shift of the received signal and T represents the time separation

between sampling instants.

The detector in equation (4.18) is now simulated. As $\beta \rightarrow \infty$, the K-distribution tends to the Gaussian distribution. As $\beta \rightarrow 0.5$, the K-distribution deviates from that of the Gaussian in the sense of having very large tails. Four different values of β=0.5, 1.0, 1.5 and 2.0 were used for performance evaluation.

The thresholds corresponding to false alarm probabilities 10^{-k}; $k = 1,2,3,4$ are obtained through the method of extreme value theory explained in Chapter 3. Once the threshold is set the detection probabilities are obtained by simulating the LOD for received vectors consisting of the sum of the signal and disturbance vectors for various signal to disturbance ratios. The value of f_D is chosen to be zero in this simulation, a worst case situation. The number of trials in the Monte Carlo simulation for obtaining detection probabilities is equal to 10,000. The performance of the LOD is compared to that of the Gaussian detector for the multivariate K-distributed clutter. The test statistic for the Gaussian detector is given by $\underline{s}^T M^{-1} \underline{r}$. Simulations could not be carried out for sample sizes larger than 128 because of the behavior of the modified Bessel functions. The modified Bessel functions are highly nonlinear and numerical difficulties arise in the evaluation of the Bessel functions for either small arguments and large orders or large arguments and small orders. For the modified Bessel function of the second kind $K_\nu(x)$, ν must not be so large compared to x such that

$$K_\nu(x) \approx \frac{\Gamma(\nu)}{2(x/2)^\nu} \tag{4.40}$$

overflows. With reference to equation (4.38) it is noticed that for $\nu = \beta - N > 128$, the value $x = 2(2\beta p)^{\frac{1}{2}}$ frequently is small enough to result in overflow. However, we notice that in equation (4.38), the test statistic has a ratio of modified Bessel functions with the order differing by one. By using the small argument approximation given in equation (4.40), the overflow problem for small arguments and large orders can be overcome. On the other hand, when the argument is large and the order is small underflow problems result. IMSL uses an iterative scheme to generate the modified Bessel functions of the second kind, starting from lower orders and building upto higher orders. The lower order Bessel functions needed to generate the higher order Bessel functions overflow for sufficiently large values of the argument. Because of this numerical difficulty analysis of performance is restricted to sample sizes that

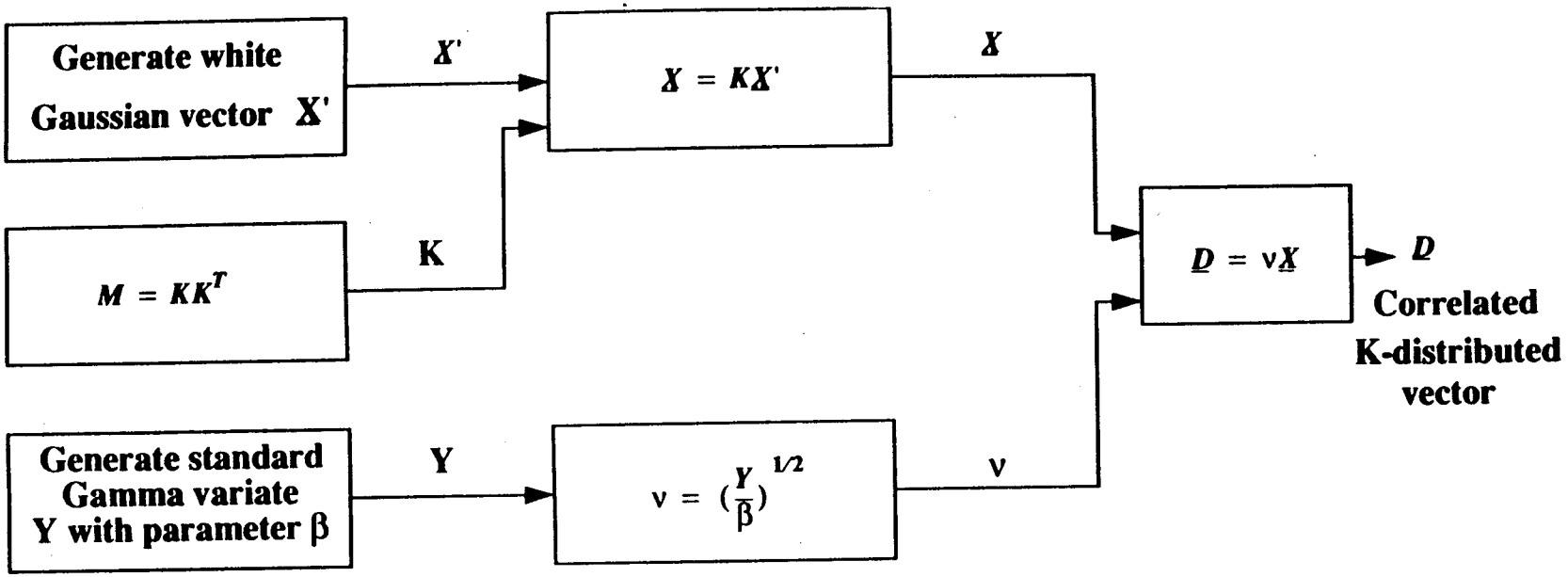


Figure 4.6: Generation scheme for the correlated multivariate K-distributed vector

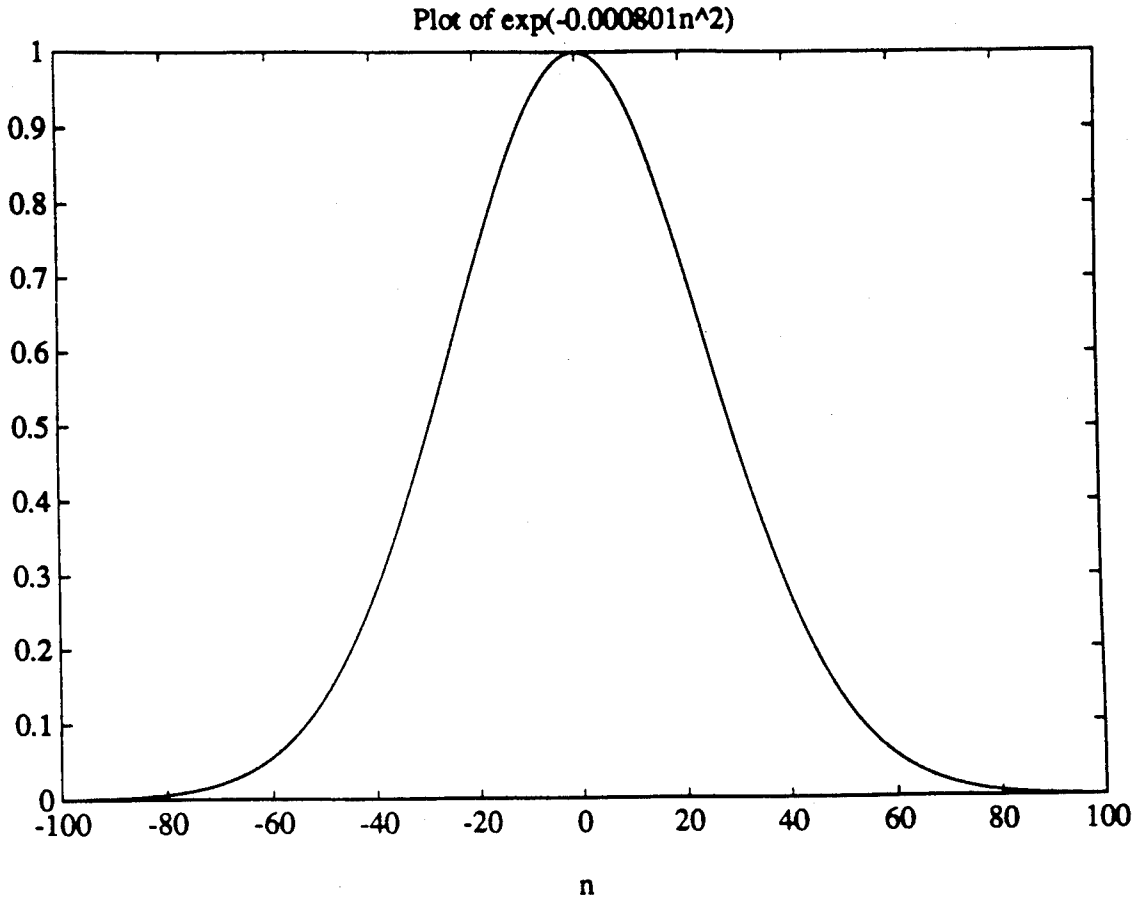


Figure 4.7: The autocorrelation function of the clutter process

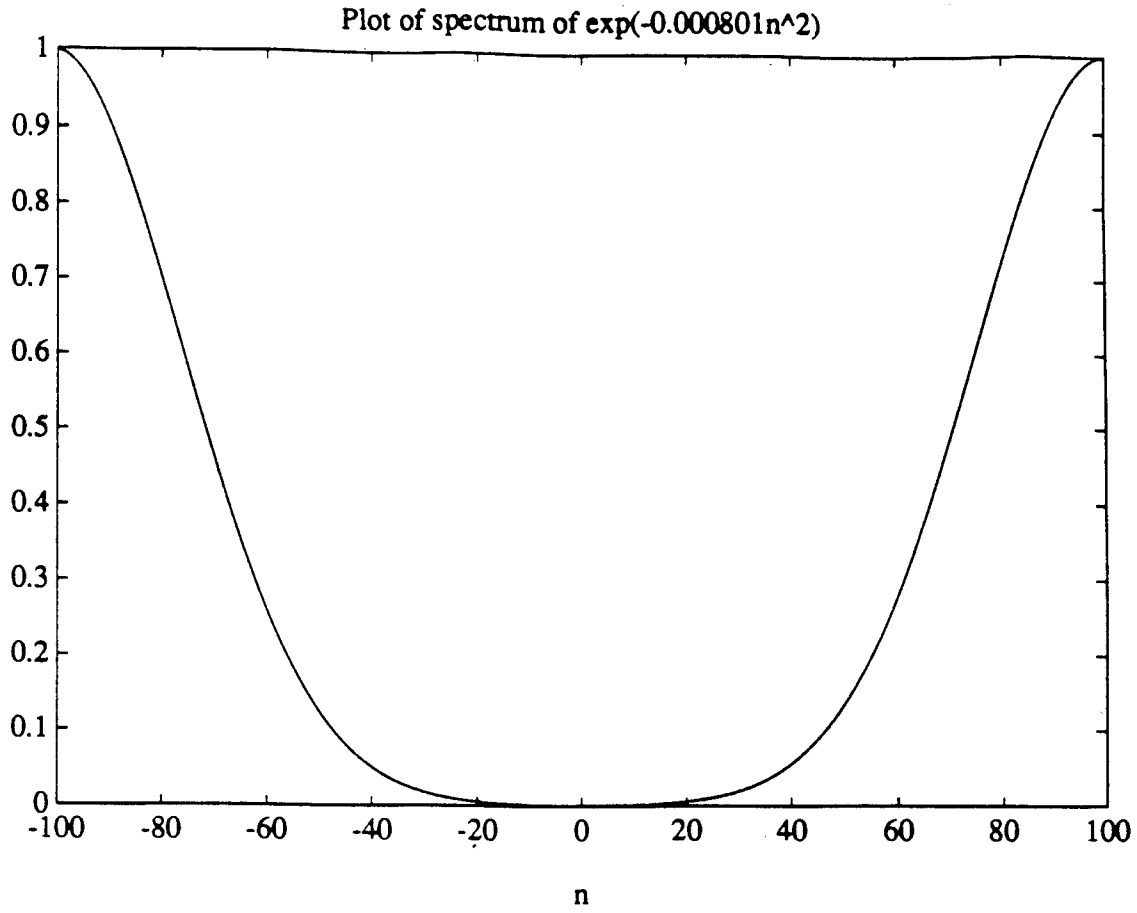


Figure 4.9: Power spectral density of the clutter process

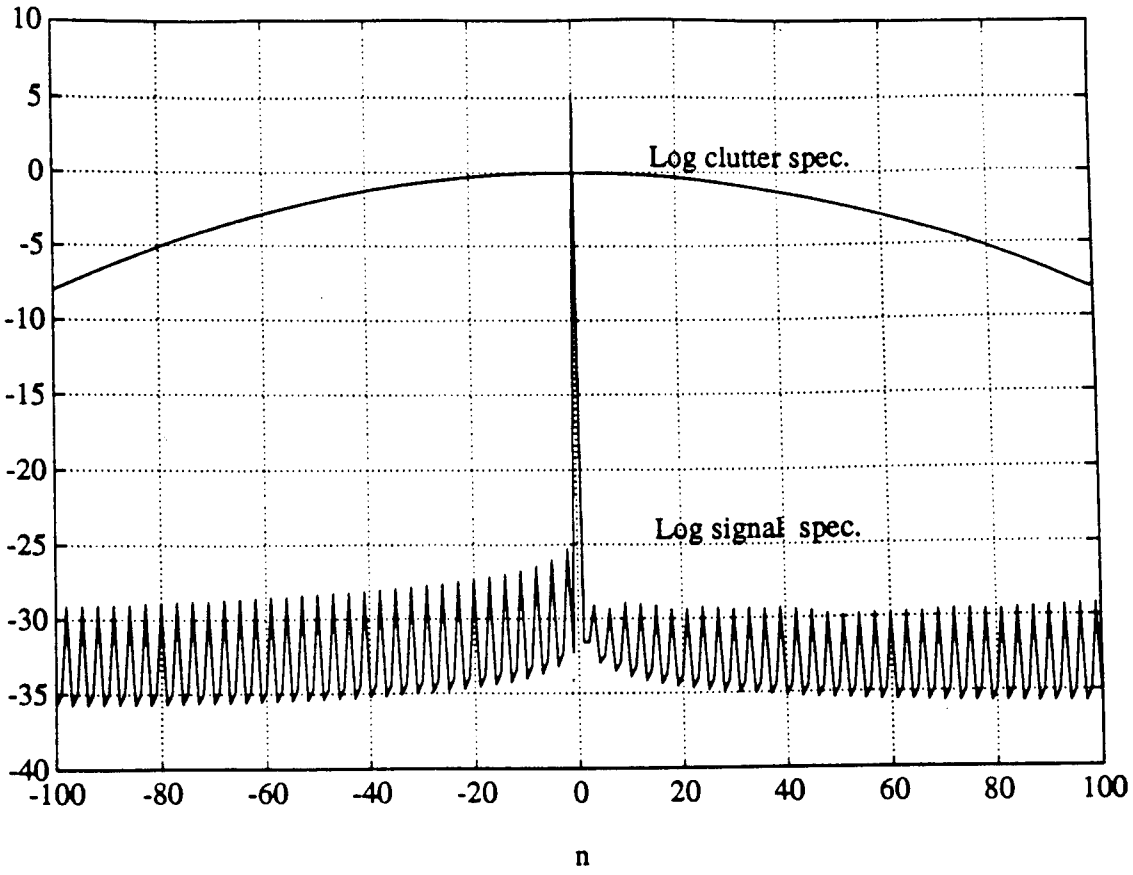


Figure 4.10: Log power spectral density of the clutter and signal processes

are smaller than 128.

4.2.3 Conclusions

When the sample size was less than 128, the probability of detection for the LOD was always less than 0.1 for false alarm probabilities less than or equal to 10^{-2}. Consequently, in this section results are tabulated for N=128. As mentioned previously, numerical difficulties made it impossible to determine performance for sample sizes greater than 128. Results are presented in Tables 4.11-4.14 for values of the shape parameter equal to 0.5, 1.0, 1.5 and 2.0 when N=128 and $P_F = 10^{-2}$. Recall that $\beta = 0.5$ corresponds to a very large tail while $\beta = 2.0$ results in a distribution that begins to approximate the Gaussian tail. Note that P_D for the LOD is relatively constant when -10 dB$\leq$SCR$<$0 dB. The best performance is achieved for $\beta = 0.5$ where P_D peaks at a value of 0.23 and the LOD performs signicantly better than the Gaussian receiver over a dynamic range of 20 dB extending from SCR equal to 0 dB to -20 dB. For $\beta = 1.0$ and 1.5, P_D peaks around 0.18 and 0.19 with a useful dynamic range of approximately 11 dB extending over the range -17 dB$\leq$SCR$\leq$-7 dB. Finally, for $\beta = 2.0$ the peak value of the LOD is 0.16 and the LOD performs better than the Gaussian receiver over a 7 dB dynamic range extending extending between -9 dB and -15 dB. The LOD for the K-distributed disturbance does not peak at values of P_D as large as those for the student-T distribution. This result agrees with that of Spaulding [18] who shows that we cannot arbitrarily say, by inspection, that a noise process that is "tremendously" non-Gaussian (i.e. the noise distribution has a very large tail) can result in "tremendous" improvement over the corresponding Gaussian or linear receiver situation. This behavior is explained by once again examining the role of the nonlinearity and the matched filter in determining the test statistic. As was the case with the student-T distribution, the nonlinearity maps large values into small values and vice versa. However, when $\beta = 0.5$, the tail of the K-distribution is much heavier than any of those encountered with the student-T distribution. In addition to more large values of the received signal due to the clutter, there are also more small values. This prevents the threshold for the K-distribution from being lowered as much as it was for the student-T distribution. As a result, if a detection is to occur, a larger boost must be generated by the matched filter when a signal is present. Unfortunately, with reference to equation (4.38), the increase due to the

SCR		LOD	GR
5 dB	P_D	0.13	0.99
1 dB	P_D	0.21	0.25
0 dB	P_D	0.21	0.16
-3 dB	P_D	0.22	0.12
-6 dB	P_D	0.22	0.09
-8 dB	P_D	0.23	0.07
-10 dB	P_D	0.23	0.05
-12 dB	P_D	0.20	0.03
-14 dB	P_D	0.17	0.02
-15 dB	P_D	0.15	0.01
-18 dB	P_D	0.12	0.01
-20 dB	P_D	0.08	0.01

Table 4.11: N=128, $P_F = 10^{-2}$, $\beta = 0.5$ SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

matched filter under hypothesis H_1 does not dominate the decrease due to the non-linearity. Consequently, not as many detections result as occurred with the student-T distribution. Finally, it is pointed out that the detection probability was much less than 0.1 for all values of the shape parameter with N=128 and $P_F = 10^{-3}$.

It should be noted that dramatically improved performance might occur when the sample size is greater than 128. Because the LOD is nonlinear, a threshold effect exists. It is not clear that N=128 is a sufficiently large sample size to get over the threshold. On the other hand, larger sample sizes cause both numerical difficulties and may not be achievable in an actual application.

In the next chapter we come up with an alternative scheme and derive a detector with enhanced performance under weak signal conditions. This detector is termed the amplitude dependent locally optimum detector. This detector is not uniformly most powerful. Thus, the thresholds for obtaining the desired false alarm probabilities and the detection probabilities are functions of the signal to clutter ratio, θ. It is shown that this detector offers a significant improvement in performance for smaller sample sizes compared to the LOD obtained on the basis of the uniformly most powerful test.

4.3 Determining Locally Optimum Detector Threshold with Real Data

Since the locally optimum detectors corresponding to SIRP multivariate density functions are nonlinear, it is not possible to evaluate the thresholds corresponding to

SCR		LOD	GR
5 dB	P_D	0.10	1.0
0 dB	P_D	0.17	0.35
-3 dB	P_D	0.17	0.33
-6 dB	P_D	0.17	0.19
-7 dB	P_D	0.18	0.15
-8 dB	P_D	0.18	0.13
-10 dB	P_D	0.18	0.04
-11 dB	P_D	0.16	0.04
-13 dB	P_D	0.14	0.04
-15 dB	P_D	0.12	0.03
-17 dB	P_D	0.10	0.03
-20 dB	P_D	0.08	0.02

Table 4.12: N=128, $P_F = 10^{-2}$, $\beta = 1.0$ SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
5 dB	P_D	0.11	1.0
0 dB	P_D	0.18	0.40
-3 dB	P_D	0.18	0.37
-6 dB	P_D	0.19	0.24
-7 dB	P_D	0.19	0.19
-8 dB	P_D	0.19	0.17
-10 dB	P_D	0.19	0.03
-12 dB	P_D	0.16	0.03
-15 dB	P_D	0.12	0.026
-17 dB	P_D	0.09	0.02

Table 4.13: N=128, $P_F = 10^{-2}$, $\beta = 1.5$ SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
5 dB	P_D	0.06	0.99
0 dB	P_D	0.14	0.44
-3 dB	P_D	0.14	0.41
-6 dB	P_D	0.15	0.25
-8 dB	P_D	0.15	0.18
-9 dB	P_D	0.16	0.14
-10 dB	P_D	0.16	0.08
-12 dB	P_D	0.14	0.05
-14 dB	P_D	0.11	0.03
-15 dB	P_D	0.09	0.02

Table 4.14: N=128, $P_F = 10^{-2}$, $\beta = 2.0$ SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

different false alarm probabilities in closed form. However, it is shown in Chapter 3 that by using extreme value theory we can model the tail behavior of the test statistic using empirically available outputs of the LOD. It is also shown that this technique yields fairly accurate thresholds for sample sizes that are orders of magnitude smaller than those required for Monte Carlo simulation. In practice, the cell false alarm probability is determined by specifying the number of false alarms allowed per scan of the surveillance volume. For example, if one false alarm per scan is specified and the surveillance volume consists of one million resolution cells, then the false alarm probability for each cell is set at 10^{-6}. The weak signal detector should be applied only to those cells for which a weak signal problem exists, as discussed in Section 1.3. The number of such cells is typically a small fraction of the total number of cells in the surveillance volume. In addition, because of the difficulty in detecting weak signals, one should allow for a few more false alarms than is the case for strong signal detection. For example, assume a surveillance volume is comprised of one million resolution cells and that 1 percent of these cells can be classified as "weak signal". Allowing for one false alarm per scan due to the "weak signal" cells, the false alarm probability for each of these cells would be 10^{-4}. If reports from the tracker are used to sort out false alarms, even larger false alarm probabilities, such as 10^{-3} might be acceptable. The false alarm probabilities can be made even larger as the number of "weak signal" cells become smaller.

In practice, reference cells centered around the test cell are used to determine the threshold for the test cell. The problem is complicated by the fact that reference cells too far removed from the test cell may not be representative due to nonhomogeneities in the clutter. As a result, the number of representative reference cells is limited. In fact, there may not be available the number of representative reference cells required by the extreme value theory. This poses a practical problem in terms of implementation of the LOD.

Fortunately, the behavior of the LOD is such that this problem can be overcome. As seen from Figures 4.1 and 4.5, the nonlinearity is such that, it transforms large values of the received observations (corresponding to large values of the quadratic form p) to small values and small values of the received observations (corresponding to small values of the quadratic form p) to large values. This implies that the tail behavior of

the LOD test statistic is governed by the body of the multivariate SIRP disturbance PDF. This is due to the fact the tail region of the test statistic corresponds to large values of the test statistic. Due to the nonlinearity, these in turn, arise from small values of the quadratic form p, which correspond to the body of the disturbance PDF.

The locally optimum detector based on the K-distributed disturbance was simulated to see if, indeed, the tail of the test statistic corresponded to disturbance values that arise from the body of the K-distributed disturbance. Since the disturbance observation space is an N-dimensional space, corresponding to N points in the observation vector, the body and tail of the disturbance is defined through the corresponding quadratic form p. If the observations are uncorrelated, then p corresponds to a hypersphere in N-dimensional space. The body of the disturbance density function, then can be defined as those points that lie within the sphere and the tail as those points that lie outside the sphere. For a SIRP disturbance the univariate density function of p can be evaluated in closed form. This enables us to evaluate the point $p = p_0$ such that $\int_0^{p_0} f_P(p)dp = \delta$, where $f_P(p)$ is the probability density function of the quadratic form p and δ is a number between 0 and 1. By choosing δ say equal to 0.98, we define the point p_0. Then we classify the simulated disturbance vectors as follows: If $p > p_0$, the vector arises from the tail of the disturbance PDF. On the other hand, if $p < p_0$, the vector arises from the body of the disturbance PDF. If the vector is correlated the treatment is similar. In this case, p corresponds to $\underline{r}^T M^{-1} \underline{r}$, where M is the covariance matrix of the disturbance. A constant value of p then corresponds to an ellipse in a N-dimensional space. A value for p_0, can be defined in the same way as was done for the uncorrelated vector situation. When the vector is uncorrelated note that M is the identity matrix.

To verify this idea simulations were carried out using the same covariance matrix as in section 4.2.2. The value of $\delta = 0.98$ was chosen. From the analytical expression for the PDF of p corresponding to the K-distributed SIRP, the value of p_0 was found to be equal to 2.15. Hence, observation vectors that resulted in values of $p > 2.15$ were assigned to to the tail of the disturbance PDF and those for which $p < 2.15$ were assigned to the body of the disturbance PDF. We now turn our attention to the test statistic. With respect to the PDF of the test statistic, the tail region is defined as the region corresponding to an area of 0.02 in the tail. From the simulations it was seen

that the small values of p mapped on to either the left tail (i.e negative) or the right tail (i.e. positive) of the test statistic PDF. Note that the output of the nonlinearity is nonnegative while the output of the matched filter can be either positive or negative. Consequently, the sign of the matched filter output determines whether small values of p map into the left or right tail. The PDF of the test statistic under the null hypothesis is symmetric with respect to the origin. Therefore, the right and the left tail were each assigned an equal area of 0.01. The threshold estimated using 10,000 computer simulations for a false alarm probability of 0.01 was 131.22. Thus, the right tail of the test statistic is defined to be those values exceeding 131.22. $10,000$ random vectors of dimension $N = 16$ were then simulated from the K-distributed disturbance. The value of the quadratic form and the test statistic were evaluated corresponding to simulated disturbance vectors. The test statistic was then sorted in ascending order. The test statistic values that exceeded 131.22 and the corresponding quadratic form values were noted. It was found that the values of p corresponding to the test statistic values greater than 131.22 were all less than 1.72. From the data it is also seen that the smaller the value of the quadratic form, the higher is the value of the test statistic. This confirms that the tail values of the test statistic arise from the body of the disturbance PDF.

To utilize the Ozturk algorithm discussed in Section 3.1.1 it is required that approximately 100 of the neighboring cells be representative of the test cell. Ozturk's algorithm can then be used to accurately approximate the body of the multivariate PDF. Once the body of the disturbance PDF is approximated accurately, it can be employed to generate the much larger sample size required to estimate the thresholds of the test statistic through extreme value theory. Good results are expected for the threshold estimation when the body of the disturbance PDF has been accurately approximated by the Ozturk algorithm because the tail of the test statistic is caused by the body of the disturbance PDF. Thus, we see that the number of reference cells required for threshold estimation is approximately 100 although we still have to generate from the approximated disturbance PDF much larger sample sizes off line in order to make use of the extereme value theory.

When the disturbance random variables arise from the student-T distributed disturbance, a similar nonlinear mapping is seen whereby the tail of the test statistic

is caused by the body of the disturbance PDF and vice versa. A procedure similar to the one followed for the K-distributed disturbance case can also be used to reduce the number of reference cells. Such a reduction in the number of required reference cells makes the implementation of these locally optimum detectors practical.

Chapter 5

Performance of the Amplitude Dependent Locally Optimum Detector

In Chapter 4 we analyzed the performance of the locally optimum detector when the underlying disturbance was modeled by the student-T and K-distributions, respectively. It was seen that the performance of the LOD was not as good as desired for the case in which the disturbance is approximated as K-distributed. In this chapter we come up with an alternate form of the the locally optimum detector which takes into account the amplitude of the weak signal. Such a detector is no longer uniformly most powerful in the sense that the thresholds for different false alarm probabilities and detection probabilities is a function of the signal to clutter ratio. We will then compare the performance of the amplitude dependent locally optimum detector (ALOD) with that of the LOD for the K-distributed and student-T disturbance models.

The uniformly most powerful test for a deterministic signal utilizes the ratio of the derivative with respect to the signal strength of the N^{th} order joint PDF under H_1 to the N^{th} order joint PDF under H_0. The limit of this ratio as the signal strength tends to zero is evaluated to obtain the test statistic for the decision rule. The amplitude dependent locally optimum detector also utilizes the ratio of the derivative with respect to the signal strength of the N^{th} order joint PDF under H_1 to the N^{th} order joint PDF under H_0. However, for the ALOD we do not evaluate this ratio as the signal strength tends to zero but leave it as a function of θ, the square root of the signal to clutter ratio. Note that θ also corresponds to the signal strength since,

as explained in Chapter 2, the variance of the clutter is taken to be unity. Because θ is unknown a priori, a bank of receivers tuned to different values of θ must be implemented. Such an approach is analogous to that of a Doppler filter bank used in Range-Doppler processing of radar signals. Instead of having a a bank of Doppler filters the ALOD employs a bank of amplitude filters. The output of each filter will be maximized when the signal amplitude is matched to the amplitude for which the filter is designed.

5.1 The Amplitude Dependent Locally Optimum Detector for the Multivariate K-Distributed Disturbance

The amplitude dependent locally optimum detector for a given value of the signal to clutter ratio, θ, is defined to be

$$\frac{\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta}}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{5.1}$$

Assuming the disturbance can be modeled by a multivariate K-distribution, $f_{\underline{D}}(\underline{r})$ under H_0 is given by equation (4.29) where $p = \underline{r}^T M^{-1}\underline{r}$. On the other hand, under H_1 the quadratic form in $f_{\underline{D}}(\underline{r}-\theta\underline{s})$ is $p_\theta = (\underline{r}-\theta\underline{s})^T M^{-1}(\underline{r}-\theta\underline{s})$ where the subscript θ is used to emphasize that we are not taking the limit as θ approaches zero. Since equation (5.1) is a ratio test, all constants can be placed in the threshold which is determined by specifying a false alarm probability. Therefore, the multiplicative constants are ignored for convenience. Hence, we will be concerned only with the terms containing the vector $\underline{R}$. Excluding constants the numerator in the ratio test is given by

$$\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta} = \frac{\partial}{\partial\theta}[(p_\theta)^{\frac{\beta-N}{2}} K_{N-\beta}[(2\beta p_\theta)^{\frac{1}{2}}]. \tag{5.2}$$

Applying the chain rule, the derivative with respect to θ can be expressed as the derivative with respect to p_θ times the derivative of p_θ with respect to θ. The derivative of p_θ with respect to θ can be derived as

$$\frac{\partial p_\theta}{\partial\theta} = (\frac{\partial}{\partial\theta}(\underline{r}-\theta\underline{s})^T M^{-1}(\underline{r}-\theta\underline{s})) = -2\underline{s}^T M^{-1}(\underline{r}-\theta\underline{s}). \tag{5.3}$$

Therefore, differentiating the right hand side of equation (5.2) with respect to θ and excluding the constant, we get

$$\begin{aligned}\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta} &= -l_\theta\left(\frac{\beta-N}{2}(p_\theta)^{\frac{\beta-N}{2}-1}K_{N-\beta}[(2\beta p_\theta)^{\frac{1}{2}}]\right. \\ &+ \left.(p_\theta)^{\frac{\beta-N}{2}}K'_{N-\beta}[(2\beta p_\theta)^{\frac{1}{2}}](\frac{\beta}{2p_\theta})^{\frac{1}{2}}\right) \qquad (5.4)\end{aligned}$$

where $l_\theta = 2\underline{s}^T M^{-1}(\underline{r}-\theta\underline{s})$, the subscript θ on l is again used to emphasize that we are not taking the limit as θ approaches zero and the prime denotes the differentiation operation with respect to the argument of the function. Equation (5.4) can be simplified further by using equations (4.36-4.37) The amplitude dependent locally optimum test can then be written as

$$T_{ALOD}(\underline{r}) = \frac{l_\theta}{(p_\theta)^{\frac{1}{2}}}(\frac{p_\theta}{p})^{\frac{\beta-N}{2}}(\frac{K_{N-\beta+1}[(2\beta p_\theta)^{\frac{1}{2}}]}{K_{N-\beta}[(2\beta p)^{\frac{1}{2}}]}). \qquad (5.5)$$

Notice that the Bessel function in the numerator is one order higher than the Bessel function in the denominator. This is significant. In the likelihood ratio test, the orders of the Bessel functions in the numerator and denominator are identical. This order difference between the Bessel function in the numerator and denominator of the ALOD makes the ALOD more sensitive to small perturbations than the classical form of the Neyman-Pearson test.

The computer simulation of performance was carried out in the same manner as described in section 4.2.2. The values of the shape parameter β chosen in this simulation are 0.5, 1.0, 1.5 and 3.0. With reference to equation (5.5), it is noticed for $\nu = N-\beta > 16$ that the value $x = (2\beta p)^{\frac{1}{2}}$ frequently assumes small enough values to result in overflow. Also, we notice in equation (5.5) that the test statistic has a ratio of modified Bessel functions with the order differing by one. Since the Bessel functions in the numerator and the denominator of the test statistic differ in both the argument and the order, the small argument approximation given in equation (4.40) cannot be used. Because of this numerical difficulty analysis of performance is restricted to sample size equal to 16.

5.1.1 Results of Computer Simulation

The results of the computer simulation are shown in Tables 5.1-5.4 for 16 complex samples and β=0.5, 1.0, 1.5 and 3.0, respectively. Performance is evaluated only for

values of SCR less than or equal to 0 dB because we are interested in the problem of weak signal detection. From the tables we observe that the best performance of the ALOD is obtained for $\beta = 0.5$. For this value of β the K-distribution has a maximum deviation from the Gaussian in the sense of having a large tail. When $\beta = 3.0$, the tail of the K-distribution closely approximates that of the Gaussian.

It is seen from the tables that the ALOD for the K-distribution significantly outperforms both the Gaussian receiver as well as the LOD. The performance of the LOD for K-distributed disturbance is tabulated in chapter 4, for N=128. Some of those results are shown again in Tables 5.1-5.4. Table 5.1 shows the detection probabilities obtained for different false alarm probabilities and signal to clutter ratios when $\beta = 0.5$. For $P_F = 10^{-2}$ the ALOD significantly outperforms the Gaussian receiver over a 30 dB dynamic range extending from 0 dB to -30 dB. The peak value of P_D equals 0.5. A similar result to that obtained for the performance of the LOD with a student-T distributed disturbance is observed here. The amplitude dependent locally optimum detector outperforms the Gaussian receiver by one order of magnitude for $P_F = 10^{-2}$, two orders of magnitude for $P_F = 10^{-3}$, three orders of magnitude for $P_F = 10^{-4}$ and four orders of magnitude for $P_F = 10^{-5}$. The useful dynamic range of the ALOD is smaller for decreasing values of P_F. The useful dynamic range is about 30 dB for false alarm probabilities equal to 10^{-2} and 10^{-3}, 25 dB for false alarm probability equal to 10^{-4} and 20 dB for a false alarm probability equal to 10^{-5}. Recall that the detection probabilities resulting from the LOD for the K-distributed disturbance were negligibly small when P_F was lower than 10^{-2}. The ALOD, on the other hand, yields significant values of P_D for P_F as low as 10^{-5}.

As the value of β is increased it is noticed from the tables that the peak value of P_D as well as the dynamic range of the receiver decreases. This arises because the tail of the K-distribution becomes closer to that of the Gaussian. In fact, when $\beta = 3.0$, the detection probabilities obtained with the ALOD for $P_F = 10^{-4}$ and 10^{-5} are negligible. Also, the useful dynamic range of the ALOD is about 15 dB for $P_F = 10^{-2}$ and less than 10 dB for $P_F = 10^{-3}$. It is interesting to note that the Gaussian receiver shows an improvement in performance for increasing values of β. The improvement, however, is only marginal.

The behavior of the amplitude dependent locally optimum detector can be un-

derstood as follows. With reference to equation (5.5), the ALOD can be factored into three terms. The first is l_θ. The second is $\frac{1}{(p_\theta)^{\frac{1}{2}}}(\frac{p_\theta}{p})^{\frac{\beta-N}{2}}$ and the third term is $\frac{K_{N-\beta+1}[(2\beta p_\theta)^{\frac{1}{2}}]}{K_{N-\beta}[(2\beta p)^{\frac{1}{2}}]}$. The term l_θ behaves identically to that of the Gaussian receiver. It yields a higher value when the signal is present than when it is absent case. The random variable p_θ, in general, tends to assume a lower value whether or not the signal is present, compared to the term p. Since $\frac{\beta-N}{2}$ is negative with magnitude greater than one, the second term in the ALOD assumes a value greater than one whether or not the signal is present. However, from the simulations, it is seen in most cases that the second factor in the ALOD assumes a larger value when the desired signal is presen than when it is not present. The modified Bessel function of the second kind, being a monotonically decreasing function assumes large values when the argument is small and small values when the argument is large. Since $p_\theta < p$ in general, the third term in the ALOD also has a value greater than one whether or not the signal is present. Once again, from the computer simulations, it is observed that the third term has a higher value in most cases when the desired signal is present. Thus, the amplitude dependent locally optimum detector for the K-distributed disturbance has three factors where each of the three factors, in general, assumes larger values under hypothesis H_1 than under hypothesis H_0. This contributes to the increased sensitivity of the ALOD to weak signals for the K-distributed disturbance to resulting in dramatically improved performance over that of the LOD with a much larger sample size. Recall that the LOD has two factors, one of which increases when the desired signal is present and the other of which decreases. The increase in the value of the test statistic of the ALOD when the signal is present is not restricted to weak signal situations. It will also hold true for strong signal situations. However, when there is a strong signal situation, the likelihood ratio test is the optimal test and the ALOD should not be used.

5.2 The Amplitude Dependent Locally Optimum Detector for the Student-T Distributed Disturbance

Assuming the disturbance can be modeled by a multivariate student-T distribution, $f_{\underline{D}}(\underline{r})$ under H_0 is given by equation (4.11) where $p = \underline{r}^T M^{-1} \underline{r}$. On the other hand, as before, under H_1 the quadratic form in $f_{\underline{D}}(\underline{r} - \theta \underline{s})$ is $p_\theta = (\underline{r} - \theta \underline{s})^T M^{-1} (\underline{r} - \theta \underline{s})$

SCR	P_F		ALOD	GR	LOD (N=128)
0 dB	10^{-2}	P_D	0.50	0.06	0.21
-5 dB	10^{-2}	P_D	0.46	0.04	
-10 dB	10^{-2}	P_D	0.40	0.02	0.23
-15 dB	10^{-2}	P_D	0.32	0.01	0.15
-20 dB	10^{-2}	P_D	0.22	0.01	0.08
-25 dB	10^{-2}	P_D	0.17	0.01	
-30 dB	10^{-2}	P_D	0.11	0.01	
0 dB	10^{-3}	P_D	0.41	0.003	
-5 dB	10^{-3}	P_D	0.39	0.001	
-10 dB	10^{-3}	P_D	0.35	0.001	
-15 dB	10^{-3}	P_D	0.27	0.001	
-20 dB	10^{-3}	P_D	0.17	0.001	
-25 dB	10^{-3}	P_D	0.13	0.001	
-30 dB	10^{-3}	P_D	0.08	0.001	
0 dB	10^{-4}	P_D	0.35	0.0003	
-5 dB	10^{-4}	P_D	0.32	0.0001	
-10 dB	10^{-4}	P_D	0.28	0.0001	
-15 dB	10^{-4}	P_D	0.23	0.0001	
-20 dB	10^{-4}	P_D	0.16	0.0001	
-25 dB	10^{-4}	P_D	0.09	0.0001	
0 dB	10^{-5}	P_D	0.30	10^{-5}	
-5 dB	10^{-5}	P_D	0.28	10^{-5}	
-10 dB	10^{-5}	P_D	0.25	10^{-5}	
-15 dB	10^{-5}	P_D	0.18	10^{-5}	
-20 dB	10^{-5}	P_D	0.10	10^{-5}	

Table 5.1: N=16, $\beta = 0.5$, SCR:Signal to Clutter Ratio, ALOD:Amplitude Dependent Locally Optimum Detector, GR:Gaussian, LOD:Locally Optimum Detector Receiver

SCR	P_F		ALOD	GR	LOD (N=128)
0 dB	10^{-2}	P_D	0.44	0.10	0.17
-5 dB	10^{-2}	P_D	0.40	0.06	
-10 dB	10^{-2}	P_D	0.33	0.03	0.18
-15 dB	10^{-2}	P_D	0.25	0.02	0.12
-20 dB	10^{-2}	P_D	0.17	0.01	0.08
-25 dB	10^{-2}	P_D	0.12	0.01	
-30 dB	10^{-2}	P_D	0.08	0.01	
0 dB	10^{-3}	P_D	0.35	0.004	
-5 dB	10^{-3}	P_D	0.30	0.001	
-10 dB	10^{-3}	P_D	0.24	0.001	
-15 dB	10^{-3}	P_D	0.18	0.001	
-20 dB	10^{-3}	P_D	0.10	0.001	
0 dB	10^{-4}	P_D	0.30	0.0005	
-5 dB	10^{-4}	P_D	0.27	0.0001	
-10 dB	10^{-4}	P_D	0.22	0.0001	
-15 dB	10^{-4}	P_D	0.11	0.0001	
0 dB	10^{-5}	P_D	0.19	10^{-5}	
-5 dB	10^{-5}	P_D	0.16	10^{-5}	
-10 dB	10^{-5}	P_D	0.11	10^{-5}	

Table 5.2: N=16, β = 1.0, SCR:Signal to Clutter Ratio, ALOD:Amplitude Dependent Locally Optimum Detector, GR:Gaussian Receiver, LOD:Locally Optimum Detector

SCR	P_F		ALOD	GR	LOD (N=128)
0 dB	10^{-2}	P_D	0.41	0.10	0.18
-5 dB	10^{-2}	P_D	0.36	0.06	
-10 dB	10^{-2}	P_D	0.27	0.02	0.19
-15 dB	10^{-2}	P_D	0.21	0.01	0.12
-20 dB	10^{-2}	P_D	0.12	0.01	
0 dB	10^{-3}	P_D	0.29	0.009	
-5 dB	10^{-3}	P_D	0.24	0.002	
-10 dB	10^{-3}	P_D	0.19	0.001	
-15 dB	10^{-3}	P_D	0.11	0.001	
0 dB	10^{-4}	P_D	0.20	0.0005	
-5 dB	10^{-4}	P_D	0.18	0.0004	
-10 dB	10^{-4}	P_D	0.16	0.0003	
-15 dB	10^{-4}	P_D	0.09	0.0002	
0 dB	10^{-5}	P_D	0.13	10^{-5}	
-5 dB	10^{-5}	P_D	0.10	10^{-5}	
-10 dB	10^{-5}	P_D	0.07	10^{-5}	

Table 5.3: N=16, β = 1.5, SCR:Signal to Clutter Ratio, ALOD:Amplitude Dependent Locally Optimum Detector, GR:Gaussian Receiver, LOD:Locally Optimum Detector

SCR	P_F		ALOD	GR	LOD (N=128)
0 dB	10^{-2}	P_D	0.36	0.18	0.14
-5 dB	10^{-2}	P_D	0.26	0.10	
-10 dB	10^{-2}	P_D	0.17	0.03	0.16
-15 dB	10^{-2}	P_D	0.11	0.01	0.09
0 dB	10^{-3}	P_D	0.18	0.06	
-5 dB	10^{-3}	P_D	0.13	0.009	
-10 dB	10^{-3}	P_D	0.06	0.004	

Table 5.4: N=16, β = 3.0, SCR:Signal to Clutter Ratio, ALOD:Amplitude Dependent Locally Optimum Detector, GR:Gaussian Receiver, LOD:Locally Optimum Detector

where the subscript θ is used to emphasize that we are not taking the limit as θ approaches zero. Since equation (5.1) is a ratio test, all constants can be placed in the threshold which is determined by specifying a false alarm probability. Therefore, the multiplicative constants are ignored for convenience. Hence, we will be concerned only with terms containing the vector $\underline{R}$. Excluding constants the numerator in the ratio test is given by

$$\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta} = \frac{\partial}{\partial\theta}\left[\frac{1}{(\beta-1+p_\theta/2)^{N+\beta}}\right]. \tag{5.6}$$

Applying the chain rule, as was done with the K-distributed disturbance differentiating with respect to θ and excluding the constant, the numerator in the ratio test is given by

$$\frac{\partial f_{\underline{D}}(\underline{r}-\theta\underline{s})}{\partial\theta} = \frac{l_\theta}{(\beta-1+p_\theta/2)^{(N+\beta+1)}} \tag{5.7}$$

where $l_\theta = 2\underline{s}^T M^{-1}(\underline{r}-\theta\underline{s})$. From the above equation the sufficient statistic for the amplitude dependent locally optimum detector for the multivariate student-T distribution can be written as

$$T_{ALOD}(\underline{r}) = \frac{l_\theta(\beta-1+p/2)^{N+\beta}}{(\beta-1+p_\theta/2)^{N+\beta+1}}. \tag{5.8}$$

Notice that the exponent in the denominator of the test statistic is one more than that in the numerator. In the likelihood ratio test the exponents in the numerator and the denominator of the test statistic are identical. But the factor l_θ appearing in the ALOD will not be present. The computer simulation of the ALOD for the multivariate student-T distributed disturbance was carried out as mentioned in section 4.1.2. The

results of the computer simulation are shown in Tables 5.5, 5.6 and 5.7.

5.2.1 Conclusions

As was the case for the K-distributed disturbance, performance is evaluated only for values of SCR less than or equal to 0 dB. Even though the ALOD may outperform the Gaussian receiver for values above 0 dB of SCR up to a certain level, the decision rule that should be used in such situations is the optimum one based on the likelihood ratio test. The value of $\beta = 1.5$ is chosen because it represents a heavy tail situation. Three different sample sizes of N=16, 32 and 64 were chosen for simulation purposes.

It is seen from the tables that the ALOD significantly outperforms the Gaussian receiver in the range of SCR values considered. The performance improvement of the ALOD over the Gaussian receiver is similar to that of the LOD discussed in Section 4.1.3. Compared to the Gaussian receiver, the ALOD has about an order of magnitude improvement in performance for $P_F = 10^{-2}$, two orders of magnitude improvement in performance for $P_F = 10^{-3}$ and three orders of magnitude improvement in performance for $P_F = 10^{-4}$. The performance improvement of the ALOD for the student-T distribution compared to that of the LOD is not as dramatic as was noticed for the K-distributed case. For the student-T distribution the performance of the locally optimum detector and the amplitude dependent locally optimum detector are actually quite close. The ALOD for the K-distribution significantly outperformed both the LOD and the Gaussian receiver. The ALOD for the student-T distribution, while significantly outperforming the Gaussian receiver in the range of SCR values considered, does not outperform the LOD significantly. In fact, comparing the detection probabilities resulting from the ALOD with that of the detection probabilities resulting from the LOD, it is noticed that the performance of the ALOD exceeds that of the LOD by less than a tenth for SCR values less than 0 dB.

With reference to equation (5.8), the ALOD can be factored into two terms. The first term is l_θ. The behavior of l_θ was explained in Section 5.1.1. The output of l_θ and the test statistic increase when the signal is present as opposed to the signal absent case. The second term is a ratio of two polynomials. In general, as discussed in Section 5.1.1, $p_\theta < p$ whether or not the signal is present. Therefore, in general, the ratio of the two polynomials is a number greater than unity. It is seen in the simulations that this ratio has a higher value under the signal present hypothesis

SCR	P_F		ALOD	GR	LOD
0 dB	10^{-2}	P_D	0.48	0.11	0.38
-3 dB	10^{-2}	P_D	0.36	0.052	0.32
-6 dB	10^{-2}	P_D	0.30	0.039	0.27
-9 dB	10^{-2}	P_D	0.19	0.024	0.15
-10 dB	10^{-2}	P_D	0.12	0.019	0.10
0 dB	10^{-3}	P_D	0.34	0.003	0.16
-3 dB	10^{-3}	P_D	0.23	0.002	0.13
-6 dB	10^{-3}	P_D	0.15	0.001	0.10
-8 dB	10^{-3}	P_D	0.10	0.001	0.08

Table 5.5: N=16, β = 1.5, SCR:Signal to Clutter Ratio, ALOD:Amplitude Dependent Locally Optimum Detector, GR:Gaussian Receiver, LOD:Locally Optimum Detector

SCR	P_F		ALOD	GR	LOD
0 dB	10^{-2}	P_D	0.48	0.18	0.46
-3 dB	10^{-2}	P_D	0.39	0.08	0.38
-6 dB	10^{-2}	P_D	0.34	0.06	0.31
-9 dB	10^{-2}	P_D	0.24	0.033	0.20
-10 dB	10^{-2}	P_D	0.15	0.026	0.13
-12 dB	10^{-2}	P_D	0.11	0.013	
0 dB	10^{-3}	P_D	0.37	0.004	0.26
-3 dB	10^{-2}	P_D	0.25	0.003	0.19
-6 dB	10^{-3}	P_D	0.16	0.002	0.14
-8 dB	10^{-3}	P_D	0.11	0.002	0.11

Table 5.6: N=32, β = 1.5, SCR:Signal to Clutter Ratio, ALOD:Amplitude Dependent Locally Optimum Detector, GR:Gaussian Receiver, LOD:Locally Optimum Detector

more often than it does under the signal absent hypothesis. This contributes to an increased sensitivity to the presence of weak signals. That explains why the ALOD outperforms the LOD. The increase in the value of the test statistic of the ALOD when the signal is present is not restricted to weak signal situations alone. The LOD also has two factors in the test statistic. However, when the signal is present, one factor increases and the other decreases. This is not a favorable situation for raising the output of the test statistic over the threshold for small signals. The LOD's nonlinearity for the student-T distribution, which is the second factor in the test statistic, is not as highly nonlinear as was observed for the K-distributed case. Thus, the performance improvement of the ALOD with respect to the LOD is not as significant.

SCR	P_F		ALOD	GR	LOD
0 dB	10^{-2}	P_D	0.59	0.30	0.55
-3 dB	10^{-2}	P_D	0.52	0.17	0.48
-6 dB	10^{-2}	P_D	0.46	0.10	0.40
-9 dB	10^{-2}	P_D	0.40	0.03	
-10 dB	10^{-2}	P_D	0.16	0.023	0.14
-12 dB	10^{-2}	P_D	0.12	0.012	0.09
0 dB	10^{-3}	P_D	0.40	0.005	0.36
-3 dB	10^{-2}	P_D	0.30	0.002	0.28
-6 dB	10^{-3}	P_D	0.23	0.002	0.22
-9 dB	10^{-3}	P_D	0.11	0.001	0.10
0 dB	10^{-4}	P_D	0.36	0.0007	0.25
-3 dB	10^{-4}	P_D	0.24	0.0002	0.19
-6 dB	10^{-4}	P_D	0.15	0.0001	0.13
-8 dB	10^{-4}	P_D	0.12	0.0001	0.11

Table 5.7: N=64, $\beta = 1.5$, SCR:Signal to Clutter Ratio, ALOD:Amplitude Dependent Locally Optimum Detector, GR:Gaussian Receiver, LOD:Locally Optimum Detector

Chapter 6

Conclusions

6.1 Summary

Conclusions and suggestions for future work are presented in this chapter. We have addressed the problem of weak signal detection in correlated multivariate non-Gaussian noise. Weak signal detectors were derived, based on the locally optimum decision rule, using the concept of Spherically Invariant Random Processes (SIRP) for modeling the radar disturbance . Locally Optimum Detectors (LOD) are useful only in the neighborhood of the point for which they are evaluated. For non-Gaussian problems the test statistic derived for the Locally Optimum Detector are nonlinear. Due to the non-Gaussian and nonlinear nature of the problem, thresholds needed to set specified false alarm probabilities cannot be obtained in closed form. The Generalized Pareto Distribution (GPD) in conjunction with the method of Extreme Value Theory was used to obtain accurate approximations to thresholds for specified false alarm probabilities. This was achieved with orders of magnitude fewer samples compared to Monte Carlo simulation. Performance analyses of the locally optimum detectors for the multivariate K-distribution and the student-T distribution were carried out by means of computer simulations. Finally, the concept of the Amplitude Dependent Locally Optimum Detector (ALOD) was introduced. For the K-distribution the ALOD was shown to greatly outperform the LOD.

The following significant contributions appear in this dissertation:

1. Under the assumption that the radar clutter can be modeled as a SIRP the canonical model for the Locally Optimum Detector was shown to be the product of the Gaussian linear receiver and a zero memory nonlinearity.

2. The computational requirements needed to set thresholds for very small false alarm probabilities were reduced by orders of magnitude using the GPD in conjunction with the method of Extreme Value Theory. Accurate thresholds were determined by introducing Ordered Samples Least Squares technique to estimate the parameters of the GPD. For example, only 20,000 samples were required for different distributions to establish thresholds corresponding to a false alarm probability equal to 10^{-6}.

3. In contrast to the available literature, where LODs are evaluated based on the assumption of an infinite sample size, performance of the LODs were obtained for finite sample sizes.

4. The new concept of the Amplitude Dependent Locally Optimum Detector (ALOD) was introduced.

5. Performance of the Amplitude Dependent Locally Optimum Detector was evaluated for finite sample sizes. The ALOD had a significant performance improvement over the LOD when the disturbance was K-distributed.

As a result of the above contributions, practical decisions can be made with respect to use of LODs. These decisions will be based on the available sample size, the desired detection and false alarm probabilities and the signal to clutter ratio.

6.2 Suggestions for Future Research

This research has led to many important results and has also raised a number of interesting questions. In particular, some of the issues arising as extensions to this work are to:

1. Compare performance of the LOD and the ALOD with that of the classical likelihood ratio test for a broad range of signal to clutter ratios and sample sizes. This will establish the conditions under which there is a need for weak signal detectors.

2. Analyze performance of the LOD and the ALOD for other multivariate probability density functions in the SIRP class, such as Weibull and Rician.

3. Establish confidence intervals for the thresholds estimated based on Extreme Value Theory.

4. Extend all of the above work to Space-Time processing.

5. Study the role of signal design in enhancing performance of the LOD.

Appendix A

Issues Related to Extreme Value Theory

A.1 Limiting Forms for the Largest Order Statistic

Let $X_1 \leq X_2 \leq ... \leq X_n$ be the ordered statistics of n random variables having a common distribution function $F(x)$. Assuming that the trials of drawing the random variables from the distribution function $F(x)$ are independent, the distribution function of the largest order statistic X_n is given by

$$\begin{aligned} P(X_n \leq x) &= P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) \\ &= F^n(x). \end{aligned} \tag{A.1}$$

When F is continuous but unknown, an asymptotic theory is developed for F in the range 0_+ to 1_- [45]. It is shown that positive sequences $\{a_n\}$ and $\{b_n\}$ exist such that

$$\lim_{n\to\infty} P(\frac{X_n - b_n}{a_n} \leq x) = \lim_{n\to\infty} P(X_n \leq a_n x + b_n) \to \Lambda(x) \tag{A.2}$$

or equivalently, by means of equation (A.1), that

$$\lim_{n\to\infty} F^n(a_n x + b_n) \to \Lambda(x). \tag{A.3}$$

Let $n = md$ in equation (A.3). d is a fixed positive constant so that as $n \to \infty$, $m \to \infty$. Using the fact that $n = md$, we can write

$$\lim_{m\to\infty} F^{md}(a_{md} x + b_{md}) = \lim_{n\to\infty} F^n(a_n x + b_n) \to \Lambda(x). \tag{A.4}$$

It is also true that

$$\lim_{m\to\infty}[F^m(a_m x + b_m)]^d = \lim_{m\to\infty} F^{md}(a_m x + b_m) \to \Lambda^d(x). \tag{A.5}$$

If equations (A.4) and (A.5) hold, then from a theorem of Hintchin [52], there exist numbers $A_d > 0$ and $B_d > 0$ such that

$$\Lambda^d(A_d x + B_d) = \Lambda(x) \tag{A.6}$$

for all integer values of d.

Solution of the above functional equation yields all the possible limiting forms for the distribution function $F^n(x)$. The constant A_d may or may not be unity. If it is unity, then the functional equation to be solved is given by

$$\Lambda^d(x + B_d) = \Lambda(x). \tag{A.7}$$

On the other hand, if A_d is not unity, the form of equation (A.6) stands and there exists a value $x_{0d} = B_d/(1 - A_d)$ such that

$$\Lambda^d(x_{0d}) = \Lambda(x_{0d}). \tag{A.8}$$

Constraining the solution to the above equation to be real and nonnegative, the solution is either $\Lambda = 0$ *or* 1. However, because $\Lambda(x)$ is a distribution function the value of Λ can be 0 only if x_{0d} is the lower endpoint at which $\Lambda(x_{0d}) = 0_+$ and Λ can be 1 only if x_{0d} is the upper end at which $\Lambda(x_{0d}) = 1_-$. Since A_d and B_d are assumed to be finite, x_{0d} must also be finite. Consequently, there is no loss in generality by assuming that the endpoint of interest is located at the origin (i.e., $x_{0d} = 0$). When $A_d \neq 1$, note that $x_{0d} = 0$ implies $B_d = 0$. As a result, the solutions for equation (A.6) fall into three cases which are given below.

$$1)\quad \Lambda^d(x + B_d) = \Lambda(x) \quad A_d = 1 \tag{A.9}$$

$$2)\quad \Lambda^d(A_d x) = \Lambda(x) \quad A_d \neq 1\ F = 0\ when\ x = 0 \tag{A.10}$$

$$3)\quad \Lambda^d(A_d x) = \Lambda(x) \quad A_d \neq 1\ F = 1\ when\ x = 0 \tag{A.11}$$

A.1.1 Case 1

Case (1) of equation (A.9) is solved as follows. Taking the logarithm, we have

$$log\ \Lambda(x) = d\ log\ \Lambda(x + B_d). \tag{A.12}$$

Multiplying through by a minus sign and taking the logarithm of both sides, we obtain

$$log[-log\ \Lambda(x)] = log\ d + log[-log\ \Lambda(x + B_d)]. \tag{A.13}$$

For simplicity, let

$$g(x) = log[-log\ \Lambda(x)]. \tag{A.14}$$

Then equation (A.13) becomes

$$g(x) = log\ d + g(x + B_d). \tag{A.15}$$

Equivalently,

$$g(x - B_d) = log\ d + g(x) \tag{A.16}$$

or

$$g(x) = g(x - B_d) - log\ d. \tag{A.17}$$

Adding equations (A.15) and (A.17), we obtain

$$g(x + B_d) + g(x - B_d) = 2g(x). \tag{A.18}$$

The above equation is valid for all x if and only if $g(x)$ is linear in x. Specifically, let

$$g(x) = kx + j \tag{A.19}$$

where j and k are constants. Then

$$g(x + B_d) = k(x + B_d) + j = g(x) - log\ d = kx + j - log\ d. \tag{A.20}$$

It follows that

$$kB_d = -log\ d \quad or \quad k = -\frac{log\ d}{B_d}. \tag{A.21}$$

Substituting equation (A.21) in equation (A.19), we see that

$$g(x) + \frac{x \log d}{B_d} = j. \tag{A.22}$$

Using equation (A.14), this result becomes

$$log[-log\ \Lambda(x)] + \frac{x \log d}{B_d} = j. \tag{A.23}$$

Thus, we have

$$log[-log\ \Lambda(x)] = -\frac{x \log d}{B_d} + j. \tag{A.24}$$

Hence, for case (1) of equation (A.9) to hold, $log[-log\ \Lambda(x)]$ must be linear in x.

We now solve for the sequence $\{B_d\}$. For this purpose, let $d = pq$ where p and q are both integers. Note that

$$\Lambda^{pq}(x + B_{pq}) = \Lambda(x). \tag{A.25}$$

From the above equation we get

$$\begin{aligned} \Lambda(x + B_{pq}) &= \Lambda^{\frac{1}{pq}}(x) \\ &= [\Lambda^{\frac{1}{p}}(x)]^{\frac{1}{q}} = [\Lambda(x + B_p]^{\frac{1}{q}} \\ &= \Lambda^{\frac{1}{q}}(x + B_p) = \Lambda((x + B_p) + B_q) = \Lambda(x + B_p + B_q). \end{aligned} \tag{A.26}$$

Equation (A.26) implies that

$$B_{pq} = B_p + B_q. \tag{A.27}$$

We now determine the functional dependence of the sequence $\{B_d\}$ on the subscript d. To emphasize this functional dependence, we rewrite equation (A.27) as

$$B(pq) = B(p) + B(q). \tag{A.28}$$

From the above equation, it is clear that the functional dependence is logarithmic. Thus, the solution for B_d is given by

$$B(d) = B_d = log\ d \tag{A.29}$$

Substituting equation (A.29) into equation (A.24) yields

$$log[-log\ \Lambda(x)] = -x + j \tag{A.30}$$

where j plays the role of a location parameter. Hence, without loss of generality, j is chosen to be zero. The above equation then simplifies to

$$log[-log\ \Lambda(x)] = -x. \tag{A.31}$$

Solution for $\Lambda(x)$ results in

$$\Lambda(x) = exp(-e^{-x}). \tag{A.32}$$

Equation (A.32) is the solution of equation (A.9) for case 1.

A.1.2 Cases 2 and 3

The solutions to Cases (2) and (3) of equation (A.10) and (A.11) are now derived. In both cases we have

$$\Lambda^d(A_d x) = \Lambda(x). \tag{A.33}$$

From equation (A.33) we get

$$log\ \Lambda(x) = d\ log\ \Lambda(A_d x). \tag{A.34}$$

Multiplying through by a minus sign and taking the logarithm of both sides, we obtain

$$log[-log\ \Lambda(x)] = log\ d + log[-log\ \Lambda(A_d x)]. \tag{A.35}$$

As in case 1, let

$$g(x) = log[-log\ \Lambda(x)]. \tag{A.36}$$

Then equation (A.35) becomes

$$g(x) = log\ d + g(A_d x). \tag{A.37}$$

Alternatively,

$$g(\frac{x}{A_d}) = log\ d + g(x) \tag{A.38}$$

or equivalently,

$$g(x) = -log\ d + g(\frac{x}{A_d}). \tag{A.39}$$

Adding equations (A.37) and (A.39) results in

$$g(A_d x) + g(\frac{x}{A_d}) = 2g(x). \tag{A.40}$$

The solution to the above equation is

$$g(x) = \pm k\ log\ x \quad for\ x > 0 \tag{A.41}$$

and

$$g(x) = \pm k\ log\ (-x) \quad for\ x < 0 \tag{A.42}$$

where k is a positive constant. Use of equation (A.36) in equations (A.41) and (A.42) yields

$$log[-log\ \Lambda(x)] = \pm k\ log\ x \quad for\ x > 0 \tag{A.43}$$

$$log[-log\ \Lambda(x)] = \pm k\ log\ (-x) \quad for\ x < 0. \tag{A.44}$$

For Case 2, $\Lambda = 0$ when $x = 0$. This implies $x = 0$ is the lower end point of $\Lambda(x)$. Hence, $\Lambda(x)$ is nonzero for $x \geq 0$. Therefore, our solution is given by equation (A.43) where we must choose the sign in front of k to be negative. Then

$$log[-log\ \Lambda(x)] = -k\ log\ x \qquad x \geq 0 \tag{A.45}$$

which results in

$$\Lambda(x) = exp(-x^{-k}) \qquad x \geq 0. \tag{A.46}$$

For case 3, $\Lambda = 1$ when $x = 0$. This implies that $x = 0$ is the upper endpoint of $\Lambda(x)$. Hence, $\Lambda(x)$ is nonzero for $x \leq 0$. Consequently, the solution is given by equation (A.44) where we choose the sign in front of k to be positive. Then

$$log[-log\ \Lambda(x)] = k\ log(-x) \qquad x \leq 0 \tag{A.47}$$

resulting in

$$\Lambda(x) = exp(-(-x)^k) \qquad x \leq 0. \tag{A.48}$$

Thus, the three possible forms for the limiting distribution $\Lambda(x)$ that arise as solutions to equation 1 are given as follows:

$$1) \qquad \Lambda(x) = exp(-e^{-x}) \tag{A.49}$$

$$2) \quad \Lambda(x) = exp(-x^{-k}) \qquad x \geq 0\ ,\ k > 0 \tag{A.50}$$

$$3) \quad \Lambda(x) = exp(-(-x)^{k}) \qquad x \leq 0\ ,\ k > 0. \tag{A.51}$$

A.2 Tails of Probability Density Functions

Equations (A.49-A.51) represent the the three possible limiting forms of the distribution function for almost all smooth and continuous probability density functions. By differentiating the three functions, we obtain analytical expressions for the limiting forms of the probability density functions. However, because of the differentiation, it should be recognized that these expressions may not be good approximations to the density functions. In practice, extreme value theory should always be applied to a distribution function, or equivalently, the area under the density function.

A.2.1 Case 1

The derivative of $\Lambda(x)$ is given by

$$H(x) = \frac{d}{dx}\Lambda(x) = exp(-e^{-x}).(-e^{-x})(-1) = e^{-x}exp(-e^{-x}) = exp(-x - e^{-x}). \tag{A.52}$$

In our application we are interested in the right tail of the probability density function. Since we have to set thresholds corresponding to small false alarm probabilities, the thresholds will be in the right tail of the probability density function. When x is very large, $x \gg e^{-x}$. Therefore, equation (A.52) can be simplified to obtain the PDF of the tail as

$$H(x) = e^{-x} \qquad x\ large. \tag{A.53}$$

A.2.2 Case 2

The derivative of $\Lambda(x)$ is given by

$$\begin{aligned} H(x) &= \frac{d}{dx}\Lambda(x) = exp(-x^{-k}).(kx^{-k-1}) \\ &= k\ exp(-x^{-k})e^{(-k-1)log\ x)} = k\ exp(-x^{-k} - (k+1)log\ x). \end{aligned} \tag{A.54}$$

When x is very large $log\ x \gg x^{-k}$. Therefore, equation (A.54) can be simplified to obtain the PDF of the tail as

$$H(x) = ke^{-(k+1)log\ x} = kx^{-(k+1)} \qquad x > 0, x\ large\ \ k > 0. \tag{A.55}$$

A.2.3 Case 3

The derivative of $\Lambda(x)$ for this case is given by

$$\begin{aligned} H(x) &= \frac{d}{dx}\Lambda(x) = exp(-(-x)^k).(k(-x)^{k-1}) \\ &= k\ exp(-(-x)^k)e^{(k-1)log\,(-x)} = k\ exp(-(-x)^k + (k-1)log\ x). \end{aligned} \tag{A.56}$$

When $-x$ is very large, $(-x)^k \gg log\ x$. Therefore, equation (A.56) can be simplified to obtain the PDF of the tail as

$$H(x) = ke^{-(-x)^k} \qquad x < 0, -x\ large\ \ k > 0. \tag{A.57}$$

A basic assumption in the above development is that successive trials are independent. This led to equation (A.1). In practice, as n becomes large, it may be difficult to ensure the independence of successive trials. To the extent that the assumption holds, the results in equations (A.49-A.51) are valid.

A.3 PDF of the r^{th} Order Statistic

Suppose that the ordered samples $X_1 \leq X_2 \leq ... \leq X_n$ are drawn from the distribution function $F(x)$. Let us further assume that the trials used to draw the samples from the distribution are independent. Consider the r^{th} order statistic X_r. Recall that $P(X_r \leq x)$ is the distribution function of X_r. This, in turn, is the probability that at least r of the $X_i's$ are less than or equal to x. Treating this as a Binomial problem, the distribution function is

$$F_{X_r}(x) = P(X_r \leq x) = \sum_{i=r}^{n} \frac{n!}{i!(n-i)!} F^i(x)[1 - F(x)]^{n-i} \tag{A.58}$$

where the i^{th} term in the summation is the binomial probability that exactly i of $X_1, X_2, ..., X_n$ are less than or equal to x. Equation (A.58) can also be represented in

the form of an integral,

$$F_{X_r}(x) = \frac{n!}{(r-1)!(n-r)!}\int_0^{F(x)} t^{r-1}(1-t)^{n-r}dt \tag{A.59}$$

which can be verified by using integration by parts in equation (A.59). The probability density function of the r^{th} order statistic is the derivative of $F_{X_r}(x)$ and is given by

$$\begin{aligned} f_{X_r}(x) = \frac{d}{dx}F_{X_r}(x) &= \frac{n!}{(r-1)!(n-r)!}\frac{d}{dx}\int_0^{F(x)} t^{r-1}(1-t)^{n-r}dt \\ &= \frac{n!}{(r-1)!(n-r)!}F^{r-1}(x)[1-F(x)]^{n-r}f(x) \end{aligned} \tag{A.60}$$

where $f(x) = \frac{d}{dx}F(x)$. Equation (A.60) represents the general form of the PDF of the r^{th} order statistic. If $F(x)$ is known, then the mean and the variance of the r^{th} order statistic can be calculated. The expected value of X_r is given by

$$E(X_r) = \frac{n!}{(r-1)!(n-r)!}\int_{-\infty}^{\infty} xF^{r-1}(x)[1-F(x)]^{n-r}f(x)dx. \tag{A.61}$$

An alternate form for the expected value of X_r can be obtained by letting $u = F(x)$. Therefore, $x = F^{-1}(u)$. The infinite limits of the integral in the above equation then become finite after the transformation. The transformed integral is

$$E(X_r) = \frac{n!}{(r-1)!(n-r)!}\int_0^1 F^{-1}(u)u^{r-1}(1-u)^{n-r}du. \tag{A.62}$$

The variance of the r^{th} order statistic is expressed as

$$Var(X_r) = E[(X_r - E(X_r))^2] = E(X_r^2) - E^2(X_r). \tag{A.63}$$

Making use of equation (A.60), $E(X_r^2)$ can be written as follows.

$$E(X_r^2) = \frac{n!}{(r-1)!(n-r)!}\int_{-\infty}^{\infty} x^2F^{r-1}(x)[1-F(x)]^{n-r}f(x)dx. \tag{A.64}$$

An alternate form for the expected value of X_r can be obtained by again letting $u = F(x)$. We then get

$$E(X_r^2) = \frac{n!}{(r-1)!(n-r)!}\int_0^1 [F^{-1}(u)]^2u^{r-1}(1-u)^{n-r}du. \tag{A.65}$$

The variance of X_r can be calculated from equations (A.62) and (A.65) when $F^{-1}(u)$ is known.

Bibliography

[1] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypothesis," *Annals of Mathematics*, pp. 25–138, 1930.

[2] J. Neyman and E. S. Pearson, "Sufficient statistics and uniformly most powerful tests of statistical hypothesis," *Annals of Mathematics*, pp. 113–137, 1930.

[3] D. Middleton, "Canonically optimum threshold detection," *Memorandum RM-4687-PR, RAND Corporation*, 1965.

[4] D. Middleton, *Statistical Communication Theory.* New York: McGraw Hill Co., 1960.

[5] D. Middleton, "Threshold detection and estimation in correlated interference," *Proceedings of the* 9^{th} *Zurich Symposium on EMC, Switzerland*, vol. EMC-91, pp. 7–12, 1991.

[6] D. Middleton, "Threshold vector field detectors," *IEEE Transactions on Electromagnetic Compatibility*, vol. **Vol 30. No. 4**, pp. 538–552, November 1988.

[7] J. H. Miller and J. Thomas, "Detectors for discrete time signals in non-gaussian noise," *IEEE Trans. on Information Theory*, vol. **IT-18, No. 2**, pp. 241–250, 1972.

[8] S. A. Kassam, *Signal Detection in Non-Gaussian Noise.* NY: Springer Verlag, 1989.

[9] I. Song and S. A. Kassam, "Locally optimum detection of signals in a generalized observation model: The known signal case," *IEEE Trans. on Information Theory*, vol. **36, No. 3**, pp. 502–515, 1990.

[10] I. Song and S. A. Kassam, "Locally optimum detection of signals in a generalized observation model: The random signal case," *IEEE Trans. on Information Theory*, vol. **36, No. 3**, pp. 516–530, 1990.

[11] J. Capon, "On the asymptotic efficiency of locally optimum detectors," *IRE Transactions on Information Theory*, pp. 67–71, 1961.

[12] E. Conte, L. Izzo, M. Longo, and L. Paura, "Asymptotically optimum radar detectors in non-rayleigh clutter," *IEE Proc.F, Commun., Radar, & Signal Process.*, vol. **134, No. 7** , pp. 667–672, 1987.

[13] V. G. Valeyev and I. Y. Aspisov, "Asymptotically effective nonparametric algorithms for detecting signals in non-gaussian interference," *Radiotekhnika i elektronika*, vol. **6**, pp. 1–7, 1989.

[14] K. R. Raveendra and R. Srinivasan, "Threshold detection of cpfsk in non-gaussian noise," *IEE Proc.F, Commun., Radar, & Signal Process.*, vol. , 1984.

[15] R. F. Ingram and R. Houle, "Performance of the optimum and several suboptimum receivers for threshold detection of known signals in additive, white, non-gaussian noise," *NUSC Technical Report*, vol. 6339, pp. 1–26, 1980.

[16] W. Gardner, "Structural characterization of locally optimum detectors in terms of locally optimum estimators and correlators," *IEEE Trans. on Information Theory*, vol. **IT-28, No. 6**, pp. 924–932, November 1982.

[17] D. H. Johnson and G. C. Orsack, "Relation of signal set choice to the performance of optimal non-gaussian detectors," *IEEE International Symposium on Information Theory*, 1990.

[18] A. D. Spaulding, "Locally optimum and suboptimum detector performance in a non-gaussian interfernce environment," *IEEE Transactions on Communications*, vol. **COM-33, No. 6**, pp. 509–517, June 1985.

[19] M. Bouvet, "Expansions of the likelihood ratio and applications," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. **ASSP-34, No. 4**, pp. 653–660, 1986.

[20] B. B. Shishkov and S. I. Penev, "Asymptotically optimum algorithms for detection of signals in a background of correlated interference and white noise," *Radiotekhnika i elektronika*, vol. **7**, pp. 1419–1424, 1988.

[21] J. W. Modestino and A. Ningo, "Detection of weak signals in non-gaussian noise," *IEEE Trans. on Information Theory*, vol. **IT-25, No. 5**, pp. 592–600, 1979.

[22] A. B. Martinez, P. F. Swaszek, and J. B. Thomas, "Locally optimum detection in multivariate non-gaussian noise," *IEEE Trans. on Information Theory*, vol. **IT-30, No. 6**, pp. 815–822, 1984.

[23] K. J. Sangston and K. Gerlach, "Results on the detection of signals in spherically invariant random noise," *NRL Report*, vol. 9202, pp. 1–12, 1989.

[24] E. Conte and M. Longo, "Characterization of radar clutter as a spherically invariant random process," *IEE Proc.F, Commun., Radar, & Signal Process.*, vol. **134** , pp. 191–197, 1987.

[25] E. Conte, M. Longo, and M. Lops, "Modelling and simulation of non-rayleigh radar clutter," *IEE Proc.F, Commun., Radar, & Signal Process.*, vol. **138** , pp. 121–130, 1991.

[26] M. Rangaswamy, D. D. Weiner, and A. Ozturk, "Spherically invariant random processes for modeling and distribution identification of non-gaussian random vectors," *IEEE Trans. on Aerospace and Electronic Systems*, vol. To Appear, pp. 111–124, January 1993.

[27] M. Rangaswamy, D. D. Weiner, and A. Ozturk, "Computer generation of correlated non-gaussian clutter for radar signal detection," *Submitted to IEEE Trans. on AES.*

[28] K. Yao, "A representation theorem and its application to spherically invariant random processes," *IEEE Trans. on Information Theory*, vol. **IT-19**, pp. 600–608, 1973.

[29] F. A. Pentini, A. Farina, and F. Zirilli, "Radar detection of targets in a located in a coherent k distributed clutter background," *IEE Proc.F, Commun., Radar, & Signal Process.*, vol. **139 No. 3**, pp. 239–245, 1992.

[30] H. V. Trees, *Detection, Estimation and Modulation Theory Part I.* New York: John Wiley & Sons., 1968.

[31] J. R. M. Hosking and J. R. Wallis, "Parameter and quantile estimation of the generalized pareto distribution," *Technometrics*, vol. **29, No. 3**, pp. 339–349, 1987.

[32] F. E. Harrel and C. E. Davis, "A new distribution free quantile estimator," *Biometrika*, vol. **Vol. 69**, pp. 635–640, 1982.

[33] A. Ozturk, "A new method for univariate and multivariate distribution identification." Submitted for publication to JASA.

[34] B. M. Hill, "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, vol. **Vol. 3**, pp. 1163–1174, 1975.

[35] J. Pickands, "Statistical inference using extreme order statistics," *The Annals of Statistics*, vol. **3**, pp. 119–131, 1975.

[36] I. Weissman, "Estimation of parameters and large quantiles based on the k largest observations," *Journal of the American Statistical Association*, vol. **Vol. 73**, pp. 812–815, 1978.

[37] D. D. Boos, "Using extreme value theory to estimate large samples," *Technometrics*, vol. **20**, pp. 33–39, 1984.

[38] B. Mandelbrot, "The pareto levy law and the distribution of income," *International Economic Review*, vol. **Vol. 1**, pp. 79–106, 1960.

[39] G. P. Patel, *Statistical Distributions in Scientific Work.* Holland: Reidel, Dordrecht, 1975.

[40] McCullough, "Continuous time processes with stable increments," *Journal of Business, The University of Chicago*, vol. **Vol. 51**, pp. 601–619, 1978.

[41] A. S. Paulson, D. C. Rutherford, and G. W. Ulseth, "Some application of stable laws to collective risk theory," *School of Management, Technical Report*, vol. Rensselaer Polytechnic Institute, 1981.

[42] P. M. Kahn, *Credibility Theory and Applications.* NY: Academic Press, 1975.

[43] M. Guida, D. Iovino, and M. Longo, "Comparitive performance analysis of some extrapolative estimators of probability tails," *IEEE Journal on Selected Areas in Communication*, vol. **Vol. 6**, pp. 76–84, 1988.

[44] SAS, *SAS User's Guide:Version 5 Edition.* Cary, North Carolina: SAS Institute Inc., 1990.

[45] T. de Oliviera, *Statistical Extremes and Applications.* Boston: D. Reidel Publishing Co., 1984.

[46] A. C. Davison and R. L. Smith, "Models for exceedances over high thresholds," *Journal of the Royal Statistical Society*, vol. **52**, pp. 393–442, 1990.

[47] W. H. DuMouchel, "Estimating the stable index α in order to measure tail thickness: A critique," *The Annals of Statistics*, vol. **11**, pp. 1019–1031, 1983.

[48] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. **7**, pp. 308–313, 1965.

[49] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products.* NY: Academic Press, 1980.

[50] E. Jakeman, "On the statistics of k-distributed noise," *Journal of Physics A.:Math. Gen.*, vol. **13**, pp. 31–48, 1980.

[51] G. N. Watson, *A Treatise on the Theory of Bessel Functions.* LONDON: Cambridge University Press, 1966.

[52] J. Galambos, *Asymptotic Theory of Extreme Order Statistics.* NY: Wiley, 1978.

Rome Laboratory

Customer Satisfaction Survey

RL-TR-_____

Please complete this survey, and mail to RL/IMPS, 26 Electronic Pky, Griffiss AFB NY 13441-4514. Your assessment and feedback regarding this technical report will allow Rome Laboratory to have a vehicle to continuously improve our methods of research, publication, and customer satisfaction. Your assistance is greatly appreciated.
Thank You

_____

_____

Organization Name:_____(Optional)

Organization POC: _____(Optional)

Address:_____

1. On a scale of 1 to 5 how would you rate the technology developed under this research?

5-Extremely Useful 1-Not Useful/Wasteful

Rating_____

Please use the space below to comment on your rating. Please suggest improvements. Use the back of this sheet if necessary.

2. Do any specific areas of the report stand out as exceptional?

Yes___ No_____

If yes, please identify the area(s), and comment on what aspects make them "stand out."

3. Do any specific areas of the report stand out as inferior?

Yes___ No___

If yes, please identify the area(s), and comment on what aspects make them "stand out."

4. Please utilize the space below to comment on any other aspects of the report. Comments on both technical content and reporting format are desired.

*U.S. GOVERNMENT PRINTING OFFICE: 1995-610-126-20053

MISSION

OF

ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

a. Conducts vigorous research, development and test programs in all applicable technologies;

b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

c. Provides a full range of technical support to Air Force Materiel Command product centers and other Air Force organizations;

d. Promotes transfer of technology to the private sector;

e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.